NASA Technical Paper 1788

Goddard Laboratory for Atmospheric Sciences Collected Reprints 1978-1979 Volume 1

MARCH 1981

NASA

NASA Technical Paper 1788

Goddard Laboratory for Atmospheric Sciences Collected Reprints 1978-1979 Volume 1

William C. Skillman, *Editor*
Earl R. Kreins, *Editor*
Goddard Space Flight Center
Greenbelt, Maryland

National Aeronautics and
Space Administration

**Scientific and Technical
Information Branch**

1981

The numbered papers in Volumes 1 and 2 of NASA Technical Paper 1788 have been accessioned as indicated below. Only the papers which have not been previously accessioned were processed as daughter items:

1. 79A48227
2. 79A32181
3. 79N20658
4. 80N16678
5. 80A39689
6. 79N20656
7. 79A35844
8. 79A33845
9. 79A35845
10. 79A29187
11. 79A29187
12. 79N20688
13.
14. 80N17638
15. 79A47689
16. 79N20633
17. 79N20660
18.
19. 80A10927
20. 79A44815
21. 78A47780
22. 79N10648
23. 78A26049
24. 79N20961
25.
26. 79A25356
27. 80A15279
28.
29. 80A26444
30. 80A26444
31. 80A10935
32. 79A50521
33. 78A23054
34. 79A23924
35. 79N13189
36.
37. 79A31871
38. 79A37698
39. 80A22480
40. 80A35540
41. 79N20645
42. 80A16184
43. 80A14265
44. 79N20654
45. 79N20650
46. 78A23607
47. 79N20644
48. 79N20574
49. 79N20639
50. 79A38006
51.
52. 79N20651
53. 79A38007
54. 78A36438
55. 81A13262
56. 79A45462
57. 79A34413
58. 80A21374
59. 79A31575
60. 79N20649
61. 80A15755
62. 78A47196
63. 79A22202
64. 78A35348
65. 78N26677
66. 78N21690
67. 79N13476
68. 79N16479
69. 79N30611
70. 79N19519
71. 80A34280
72. 78N20575
73. 79A48230
74. 80N12536
75. 79N16473
76. 78N33284
77. 80N19591
78. 80N12534
79. 78A35349
80. 79N10502
81. 79N23561
82.
83.
84. 78N20576
85. 79N28649
86. 80A20247
87. 80N19588
88.
89. 79N33725
90.
91. 79A22206
92. 80A15556
93. 79A53581
94.
95. 78N20577
96. 79N23478
97.
98. 78N16410
99.
100. 79A20134
101. 77N19564
102. 77N19563
103. 79A53565
104. 79N20672
105. 80N16403
106. 81A11971
107. 81A11970
108. 79N27635
109. 79N20689
110. 80A20194
111. 79A17829
112. 78A43280
113. 79A47678
114. 80N16402
115. 79N16494
116. 79N33531
117. 79N20674
118. 79N20676
119. 78A50493
120. 79A43934
121. 79A16717
122. 79A23668
123. 80A13586
124. 79A47192
125. 80A12813
126.

FOREWORD

The Goddard Laboratory for Atmospheric Sciences (GLAS) was established in the Fall of 1977 at the Goddard Space Flight Center to provide a balanced program of basic and applied research dedicated to advancing the understanding of the complete atmospheric, hydrospheric, and cryospheric systems. The Laboratory is an arm of the Applications Directorate of the Goddard Space Flight Center and its goals reflect those of the NASA Office of Space and Terrestrial Applications, which has overall programmatic responsibility.

This is the first edition of Collected Reprints, presented here in two volumes, comprising material published by members of the Laboratory between January 1, 1978 and December 31, 1979. Future issues will cover just one calendar year. To avoid unnecessary duplication, only the abstracts or introductions of NASA reports and conference proceedings are included.

During this period, the Laboratory consisted of the Modeling and Simulation Facility, Upper Atmosphere Branch, Troposphere Branch, Hydrospheric Sciences Branch, Climate and Radiations Branch, and Special Projects Office.

The main purpose of this publication is to aid in communicating the research efforts of GLAS to the scientific community. It is recognized that many people will have seen the papers as they appeared in journals and other publications. However, many researchers, especially those in sister disciplines, may not have had ready access to such publications. Thus, it is hoped that this collection will provide a welcome and ready reference not only to our colleagues conducting parallel research in other agencies and academic institutions, but also to those in related areas.

The reprints have been grouped according to the internal structure of the Laboratory:

Special Projects Office—management of special project activities, such as the Global Atmospheric Research Program (GARP), Operational Satellite Improvement Program (OSIP), Ozone Processing Team (OPT), and applications development.

Modeling and Simulation Facility—develops general circulation models to describe atmospheric behavior and simulate observational system capabilities. Conducts a research program of computer experiments on the application of space-borne data for global weather, climate, and ocean predictability studies.

Upper Atmosphere Branch—research on the structure, composition, energetics, and radiative properties of the troposphere, stratosphere, and mesosphere and investigations of sun-weather/climate relationships.

Troposphere Branch—research on storms, boundary layer processes, and local, medium, and regional scale processes.

Hydrospheric Sciences Branch--research related to developing an improved understanding of space and applications of such to the hydrological, oceanographic, and cryospheric sciences.

Climate and Radiations Branch--empirical and theoretical studies of climatic processes and relationships, steady-state climate models, climate sensitivity to parameter variations, earth radiative energy budget, and effects of the atmosphere on remote sensing.

CONTENTS

CONTENTS (continued)

CONTENTS (continued)

III. UPPER ATMOSPHERE BRANCH

CONTENTS (continued)

IV. TROPOSPHERE BRANCH

CONTENTS (continued)

CONTENTS (continued)

CONTENTS (continued)

CONTENTS (continued)

CONTENTS (continued)

CONTENTS (continued)

CONTENTS (continued)

CONTENTS (continued)

I. SPECIAL PROJECTS OFFICE

Paper 1

[Reprinted from BULLETIN OF THE AMERICAN METEOROLOGICAL SOCIETY, Vol. 60, No. 7, July 1979]
Printed in U. S. A.

The Data Systems Tests: The Final Phase

James R. Greaves

Laboratory for Atmospheric Sciences, Goddard Space Flight Center
Greenbelt, Md. 20771

1. Introduction

The United States has conducted a series of Data Systems Tests (DSTs) as a precursor to its participation in FGGE, the Global Weather Experiment. This paper briefly describes the impact those tests have had on the FGGE observing system and on the data management plans. In particular, the final phase of the DST programs is described, wherein a number of investigators have been selected to work with the DST data sets in research studies directed toward the GARP objectives. By drawing attention to these investigations at a relatively early stage, we are taking an important first step in providing feedback to the potential FGGE research community.

2. The Data Systems Tests

One of the important lessons learned from the earlier GARP field experiments such as the Barbados Oceanographic and Meteorological Experiment (BOMEX) and the GARP Atlantic Tropical Experiment (GATE) was the need for limited but realistic simulations prior to attempting the larger experiment. It was in this light that the concept of conducting a series of Data Systems Tests as a precursor to FGGE was developed.

The GARP Project at NASA was charged with the task of implementing the DST tests. The approach was to use current operational satellites, conventional sources of meterorological data, and data from research satellites to approximate the elements of the proposed observing system for FGGE. The goal was to provide an early look at potential problems associated with the data acquisition, data management, and data utilization aspects of the Global Weather Experiment itself.

Beginning in late 1973 and continuing into 1976, NASA, in collaboration with NOAA, NCAR, the University of Wisconsin, and others, conducted a series of six Data Systems Tests (Fig. 1). Of these, the first four were more limited tests of individual subsystems. The final two tests, DST-5 and -6, consisted of comprehensive tests of the entire system. They were of 60 days duration each and were conducted under summer and winter conditions, anticipating the two Special Observing Periods of FGGE.

3. The impact of the DSTs

As a result of the DST experience, significant changes have been made in both the planned observing system for FGGE and the procedures for managing the data

0003-0007/79/070791-04$05.00

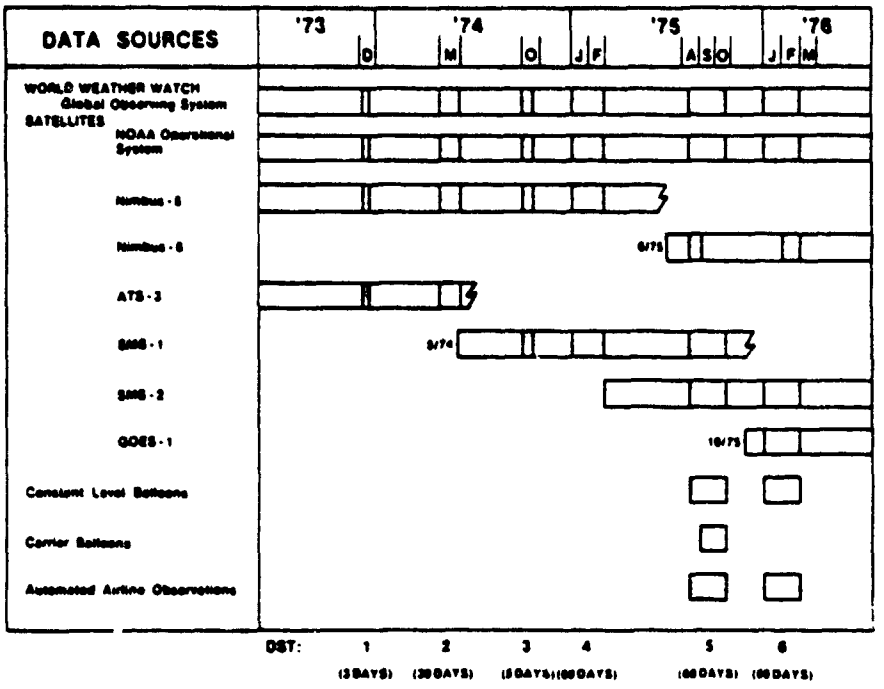


FIG. 1. During the Data Systems Tests, currently available observing systems were used to simulate the data to be acquired during the Global Weather Experiment. The final two DST periods were full end-to-end systems tests yielding comprehensive sets of global meteorological data. It is these data sets that will be used in support of the DST Research Program.

flow. Some typical examples of these changes are given in the following sections.

a. Data acquisition

1) The concept of using a Carrier Balloon System carrying deployable dropsondes to obtain upper-level wind data was thoroughly evaluated and subsequently dropped in favor of the aircraft dropwindsonde program and increased use of dedicated ships.

2) The Aircraft Integrated Data System (AIDS), which provides automated reports from commercial aircraft, was shown to be a valuable source of upper-air data during both GATE and the DSTs.

3) In order to provide flight level data in near real time, the Aircraft to Satellite Data Relay (ASDAR) system was developed by NASA and is currently being used in the Global Weather Experiment. This system assembles the same meteorological data as AIDS, but the data are transmitted hourly via geosynchronous satellites to ground stations, rather than being stored in the aircraft on tape cassettes.

4) The performance of the Nimbus-6 sounding system during the DSTs gave added support to the decision to employ that system on the TIROS-N and NOAA-A satellites for FGGE. The fact that sounding retrieval schemes had already been worked out for the Nimbus sounder greatly accelerated the availability of the TIROS-N data-processing software.

b. Data management

1) The need for an operational readiness evaluation of the data flow prior to the start of FGGE was demonstrated. The FGGE data management plan, which allowed a full year for testing national and international interfaces before the experiment began, is based largely on the experience gained in processing the DST data.

2) The need for a greatly improved monitoring of the observing and telecommunications systems was demonstrated. As a result of its DST experiences, the United States has been a prime mover in developing a World Weather Watch Monitoring Plan.

3) Early attempts to use the DST data revealed a number of deficiencies in the data processing and quality control procedures employed during the experiments. Because the DST experiments were designed to simulate the FGGE environment, a

special effort is being made to "edit and enhance" the U.S. operational soundings and cloud motion winds during the Special Observing Periods of FGGE. A joint NASA/NOAA program has been established to implement this effort. (See the GARP Topics article in the January 1979 BULLETIN.)

4. The research phase

The final two DST periods produced the most comprehensive sets of global meteorological data yet assembled. In terms of completeness, they will not be surpassed until the FGGE data become available. A number of studies using these data to determine the impact of satellite soundings on weather forecasting have already taken place. (Halem *et al.*, 1978; Desmarais *et al.*, 1978; Miyakoda *et al.*, 1977; Kelly *et al.*, 1978; and Miller and Hayden, 1978.) While there are differing opinions as to the degree of the impact, it is agreed that when properly utilized, the satellite sounding data do have a beneficial impact on our weather forecasting ability.

The NASA impact test results (Halem *et al.*, 1978) point to two major areas, which, if improved, can bring about larger forecast impacts from satellite sounder systems:

1) Technology must be developed to improve the accuracy and vertical resolution of the sounder temperature profiles themselves. DST satellite soundings have rms errors of approximately 2–2.5°C when compared with colocated radiosonde profiles. While the deficiencies in the quality of the data can be partially compensated for by special processing and assimilation methods, the need for more accurate temperature profiles under all conditions is still the most important requirement.

2) Numerical prediction models themselves must be improved to make better use of the sounding data. In order for these asynoptic data sources to contribute to more accurate forecasts, it is necessary that the model be able to convey information accurately over extended distances and periods of time from one region of the globe to another. Improvements in model forecasts can only proceed by a judicious combination of higher grid resolutions, more accurate numerical integration methods, and better representations of atmospheric processes.

As the final phase in the DST program, a series of research studies utilizing the DST data sets and focused on the problem areas outlined above has recently been initiated. Brief summaries of the selected DST investigations are listed below as a first step in providing feedback from these studies to the FGGE research community.

1) *Satellite Data Assimilation in Numerical Forecast Models.* The Principal Investigator is Yoshi K. Sasaki of the University of Oklahoma.

The objective is to develop new statistical analysis techniques for combining satellite data with conventional data to improve numerical analyses and forecasts. A barotropic "noise freezing" methodology (a technique to reduce the phase speed of gravity waves without resorting to damaging dissipative devices) will be adapted to baroclinic models. The impact of baroclinic noise freezing will be tested using the UCLA general circulation model, and generalized optimum interpolation techniques to accommodate the special characteristics of satellite data will be developed.

2) *Resolution Studies and Energy Diagnostics of Global Data Sets.* The Principal Investigators are Donald R. Johnson and Robert G. Gallimore, Jr. of the University of Wisconsin.

The objective is to assess the need for, and impact of, DST nonconventional data (including satellite information) in data-sparse regions through the use of diagnostics of available potential energy. An evaluation of the impact of higher resolution data will be made by comparing energy statistics computed from conventional and augmented DST data sets. An assessment will be made regarding the sensitivity of the Goddard Laboratory for Atmospheric Sciences General Circulation Model (GLAS GCM) to the augmented DST data.

3) *A Study On The Meteorological Significance of Cloud Tracked Winds During DST-5 and DST-6.* The Principal Investigator is Verner E. Suomi of the University of Wisconsin.

The objective is to study the limitations of cloud wind data in terms of accuracy, data sampling, and data density, and how these factors relate to the meteorological significance of the data sets. Comparisons will be made between the NOAA/NMC and NASA/GLAS Level III wind fields and the Level II wind fields produced by Wisconsin, NOAA/NESS, aircraft data, and radiosondes. Studies will be performed on the representation of synoptic scale features (such as fronts, jets, waves, cyclone centers, etc.) by both the Level III analyses and the Level II data.

4) *Global Atmospheric Research Program—Application of Global Data Sets.* The Principal Investigator is Takio Murakami of the University of Hawaii.

The objective is to use DST-5 and -6 Level III data sets for studies of regional energetics of the Asiatic summer and winter monsoon circulations, and of wave-wave and wave-zonal mean flow interactions for standing and transient waves. The validity of the preliminary results obtained from recent studies of regional energetics and mean flow interactions will be tested using the superior DST

data sets. Where possible, the research will be extended to additional problem areas that could not be investigated in the earlier work due to limitations of the available data.

5) *Evaluation of GARP Level III Data Sets for DST-5 and DST-6.* The Principal Investigator is James C. Sadler of the University of Hawaii.

The objective is to evaluate the quality of DST-5 and DST-6 Level III data sets for the global tropics and to provide a common standard for comparing NOAA/NMC and NASA/GLAS analysis and forecasting schemes. Subjective kinematic analysis techniques will be used to generate subjective Level III analyses at two levels. The major differences between the objectively and subjectively gridded analyses will be discussed, including their causes and available or anticipated remedies.

6) *A Study of Oceanic and Atmospheric Forcing Using GARP DSTs.* The Principal Investigators are Eric B. Kraus and Marvin A. Geller of the University of Miami.

The objectives are to study the nature of the variability of ocean forcing and to investigate the geographic distribution of tropospheric diabatic heating and its component parts of net radiational heating, sensible heat flux, and latent heat flux. The Level III surface data sets will be used to compute 6 h and 12 h grid values of ocean forcing parameters in order to determine the error introduced by the expression of these parameters in terms of averages of bulk variables. Level II data and existing parameterization formula will be used to determine the component parts of diabatic heating.

7) *An Improved Nimbus-6 Sounding Data Set For Experiments With Numerical Forecast Models.* The Principal Investigators are William L. Smith of NOAA/NESS and Verner E. Suomi of the University of Wisconsin.

The objective is to produce a high quality sounding data set from Nimbus-6 High Resolution Infrared Spectrometer (HIRS) and Scanning Microwave Spectrometer (SCAMS) radiance data, which possesses the highest achievable spatial resolution. The Nimbus-6 data will be reprocessed in a man-machine interactive manner using the McIDAS system at the University of Wisconsin.

8) *Linear and Nonlinear Interactions of the Large-Scale Atmospheric Waves.* The Principal Investigator is S. K. Kao of the University of Utah.

The objective is to investigate how the kinetic and available potential energies are transferred, converted, and dissipated in large-scale atmospheric waves of different characteristics, including the linear and nonlinear contributions to the generation and maintenance of such waves. The power- and cross-spectra of meteorological quantities will be computed in the wave number-frequency domain and the kinetics and thermal energies associated with atmospheric waves of various wavelengths moving at various phase velocities will be analyzed.

9) *The Effect of Global Data Sets on Predictability Degradation in the Southern Hemisphere.* The Principal Investigator is I. M. Navon of the Council for Scientific and Industrial Research, Pretoria, South Africa.

The objective is to use DST data to test the hypothesis that an expanded southern hemisphere data base will lead to substantial improvements in forecasts. A series of 4-day forecasts will be generated with and without the DST data sets using an existing Primitive Equations (PE) hemispheric model.

5. Conclusion

It is hoped that the investigations outlined above will serve as a precursor to the research phase of FGGE in the same way that the data acquisition and data management aspects of the DSTs have had a significant impact on their FGGE counterparts. They should provide an early opportunity for potential FGGE researchers to address problems related to the GARP objectives and thereby accelerate the payoff to be derived from FGGE itself.

References

Desmarais, A., S. Tracton, R. McPherson, and R. Van Haaren, 1978: The NMC Report on the Data Systems Test. *Final Rept., NASA Contract S-70252-AG*, NOAA/NMC, Camp Springs, Md., 323 pp.

Halem, M., M. Ghil, R. Atlas, J. Susskind, and W. Quirk, 1978: The GISS Sounding Temperature Impact Test. *NASA Tech. Memo. 78063*, NASA Goddard Space Flight Center, Greenbelt, Md., 421 pp.

Kelly, G., G. Mills, and W. Smith, 1978: Impact of Nimbus-6 Temperature Soundings on Australian Region Forecasts. *Bull. Am. Meteorol. Soc.* 59, 393–405.

Miller, A., and C. Hayden, 1978: The Impact of Satellite-Derived Temperature Profiles on the Energetics of NMC Analyses and Forecasts during the August 1975 Data Systems Test. *Mon. Wea. Rev.*, 106, 390–398.

Miyakoda, K., T. Batige, W. Davis, R. Hovaner, R. Jusen, F. Pratte, J. Ryan, and L. Umscheid, 1977: Preliminary Report on Satellite Sounder Impact Test at GFDL, 9 pp. (Unpublished Manuscript.)

Paper 2

[Reprinted from BULLETIN OF THE AMERICAN METEOROLOGICAL SOCIETY, Vol. 60, No. 2, February 1979]
Printed in U. S. A.

A "Special Effort" to Provide Improved Sounding and Cloud-Motion Wind Data for FGGE

James R. Greaves

Laboratory for Atmospheric Sciences, Goddard Space Flight Center, Greenbelt, Md. 20771

Geoffrey DiMego

National Meteorological Center, NOAA, Washington, D.C. 20233

William L. Smith

National Environmental Satellite Service, NOAA, Washington, D.C. 20233

Verner E. Suomi

Space Science and Engineering Center, University of Wisconsin, Madison, Wis. 53706

1. Introduction

Beginning in late 1973 and continuing into 1976, NASA, in collaboration with NOAA, conducted a series of Data Systems Tests (DST), which were intended to be a precursor to U.S. participation in FGGE—The Global Weather Experiment. These tests proved to be extremely useful in evaluating the suitability of the projected FGGE observing systems, the data processing and distribution facilities, and our ability to use the resultant data in analysis and forecast models. The final two tests of the DST series produced the most complete sets of global meteorological data ever assembled.

Early attempts by the modeling community to use the DST data in achieving improved numerical forecasts were disappointing. We have since learned that part of the problem stems from the existence of a number of deficiencies in the processing and quality control procedures employed during the DST experiments. Contributing to these deficiencies was the fact that certain of the key data sets were produced in an isolated environment. That is, no ancillary data were used to quality check or guide the derivation of the final product. This was particularly true of the high density cloud motion winds and the research satellite soundings. In the case of the winds, the single most important deficiency was in the accuracy of the assigned cloud heights. Little effort was made to adjust the cloud motion heights using ancillary data such as satellite temperature soundings or forecast wind fields. An added difficulty in the case of the sounding data was that they were horizontally averaged so as to be more compatible with the GARP grid size. This, together with the inherent lack of verti-

0008-0007/79/020124-04$05.00

cal resolution of satellite soundings, contributed to the tendency of the analyses based on satellite sounding data to exhibit too little variance in the thermal structure of the atmosphere. This led in turn to an underestimation of the amplitudes of weather systems.

As they were originally set up, the DST experiments were *designed* to simulate the data acquisition and processing environment of the Global Weather Experiment. Unfortunately, by the time the deficiencies in the data processing and quality control procedures were brought to light, it was too late to substantially alter existing plans for the operational processing of satellite data during FGGE. Thus, it became clear that FGGE itself would suffer from the same deficiencies unless some sort of special effort was made to circumvent these difficulties.

In January 1978, a panel of experts on quality control for FGGE recommended that a special effort be made to "enhance and edit" the satellite soundings and cloud motion winds. The panel, which consisted of both data producers and data users, stated that such an effort would be crucial for achieving the full impact of these observing systems. The panel recommended that an extensive effort be made to improve the cloud height determination, that the sounding data be processed with the highest possible horizontal resolution (particularly in meteorologically active areas), that data gaps be filled where possible, and that quality indicators be provided with the final data sets.

2. System Description

In February 1978, representatives of NASA, NOAA, and the University of Wisconsin met at NASA's Goddard Space Flight Center to discuss the planning and implementation of the recommended special effort. A one-pass system was decided upon, wherein the initial editing and selection of areas for data enhancement would be conducted by the NOAA National Meteorological Center (NMC), the sounding and wind enhancement would be carried out under the direction of the GARP Project at Goddard, and the final editing and quality assessment of the enhanced data by NOAA/NMC.

A diagram outlining the proposed implementation plan is presented in Fig. 1. Man-computer Interactive Data Access System (McIDAS) terminals developed by the University of Wisconsin will be used both for editing operations at NOAA/NMC and for data enhancement at NASA/Goddard. (McIDAS is a highly sophisticated hardware/software system developed for accessing, processing, and displaying data from high volume data sources such as meteorological satellite imaging systems.) The McIDAS operating system and applications software will be installed in the Amdahl computer system at Goddard's Laboratory for Atmospheric Sciences. The Goddard terminal will be colocated with the computer. The NMC terminal will be tied via land line to the Amdahl computer for direct access to the FGGE data sets. A third terminal will be available at the University of Wisconsin for software development and system maintenance.

The appropriate FGGE data sets will be acquired from NOAA and other sources. After the editing and enhancement procedures have been completed, the resultant data products will be forwarded to the World Data Centers and to the FGGE Level IIIb analysis

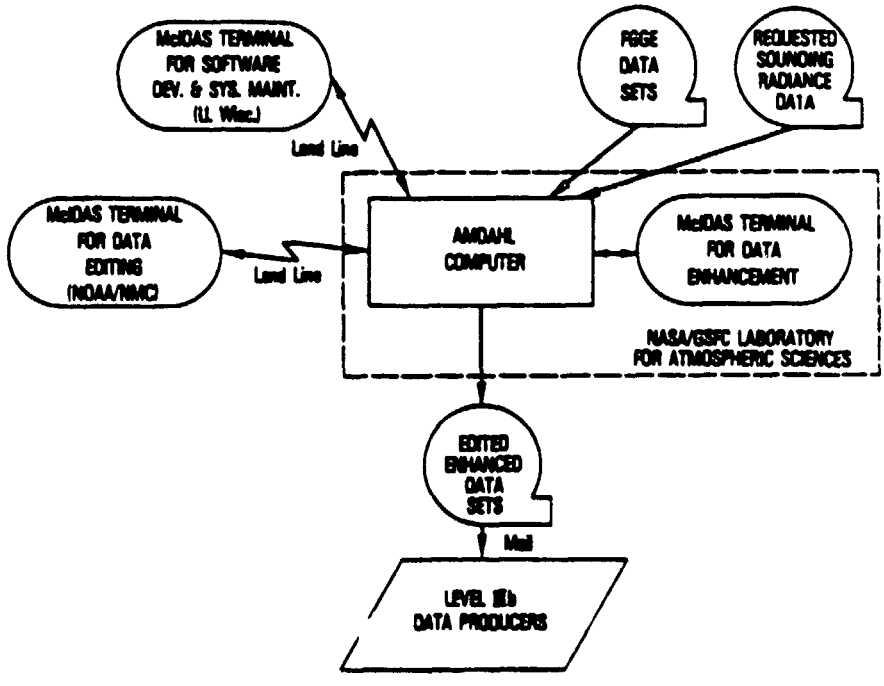


FIG. 1. Diagram of proposed implementation plan for special effort to provide improved sounding and cloud-motion wind data for FGGE.

centers as a supplement to the official FGGE data sets. (Level IIIb data are the final grid point analyses to be generated on a delayed basis from the full FGGE data sets. There are two official Level IIIb producers: NOAA's Geophysical Fluid Dynamics Laboratory and the European Center for Medium Range Weather Forecasts.)

The development of a man-interactive capability on the Amdahl at Goddard, installation of a McIDAS terminal at NMC, and the provision of training to NASA and NOAA scientists participating in the special effort will be the responsibility of the University of Wisconsin. The development of software and techniques to ingest the FGGE data, display them, edit erroneous data, and select meteorologically active regions or regions with data deficiencies will be the responsibility of both NOAA/NMC and NASA/Goddard.

3. Implementation

The special effort will be conducted only for the two Special Observing Periods of FGGE (January/February 1979 and May/June 1979). Its primary objective will be the production of edited, enhanced, and high quality sounding and cloud motion wind sets. Neither the editing nor the enhancement efforts will be compromised in quality in the interest of satisfying particular time constraints, although every effort will be made to have the data available in time for use by the Level IIIb analysis centers (nominally six months after data acquisition).

While a review of potential data problems will begin much earlier, the formal editing task will not begin until at least three months after the data observation date so as to allow the use of as broad a data base as possible. Data input will include the following:

1) global operational data received by NOAA/NMC (includes non-U.S. satellite and conventional data);
2) operational satellite sounding and wind data sets generated by the NOAA National Environmental Satellite Service (NESS)
3) Wisconsin wind data at the FGGE scale from the Indian Ocean satellite;
4) Wisconsin tropical mesoscale wind data from all three U.S. geostationary satellites;
5) special aircraft dropwindsonde data;
6) commercial aircraft data.

The data and derived products will be displayed via a McIDAS terminal to an experienced meteorologist at NMC for subjective evaluation. The data will be evaluated on the basis of synoptic reasonableness, vertical, horizontal, and temporal consistency, and consistency with other data (e.g., satellite data with nonsatellite data, satellite winds with satellite soundings, etc.). Primary emphasis will be placed on the soundings and cloud motion winds obtained from U.S. satellite systems. Quality flags will be assigned to the reviewed data. These quality indicators will be supplied in addition to those already provided by the data producer. None of the original data will be changed or discarded, only graded as to quality. An attempt will be made to document the reasons behind any particularly high or low quality assessments. Regions of special interest such as meteorologically active areas, data voids, and areas of questionable data quality will be identified as areas where enhancement efforts should be concentrated.

The sounding data enhancement process at Goddard will consist of supplementing the operational satellite sounding data set with higher resolution soundings in meteorologically active regions, and with new soundings in data voids or areas of questionable quality. Upon selection of the regions of interest, the operator will access, via the McIDAS terminal, the raw radiance data for use in producing new high resolution soundings. These new soundings will be generated using new retrieval schemes developed by NOAA/NESS in conjunction with man-interactive processing techniques.

The cloud motion wind enhancement process will consist of the rederivation of cloud heights in selected areas using new techniques developed by NESS and the University of Wisconsin that make use of sounding radiance data. In areas of significant data voids, it may be possible for Wisconsin to generate additional winds at some prescribed density. For both winds and soundings, the enhanced data sets will be identified as such, appended to the original operational products, and turned over to NMC for final editing.

At NMC a synoptician will then evaluate these new data and assign coded quality indicators that will describe the validity of various observations. In addition, NMC will gather information on the basic characteristics and systematic errors of the data, together with descriptions of the nature and possible source of the errors. After this last editing step, the final data package, including the edited and enhanced soundings and cloud motion winds as well as the edited portions of the original operational products, will be forwarded by NASA to the World Data Centers and to the FGGE Level IIIb centers.

4. Summary and Conclusion

The United States pre-FGGE experiment known as the Data Systems Tests clearly indicated a need for a special effort to edit data and enhance the quality of satellite soundings and cloud motion winds during FGGE. A panel of experts on FGGE quality control has proposed that the U.S. make a special effort to edit and enhance the satellite soundings and cloud motion winds obtained during the Global Weather Experiment. In response to that request, a formal program to carry out the editing and enhancement tasks has been established jointly by NASA, NOAA, and the University of Wisconsin. This program will be funded by the GARP Project at NASA and by the U.S. FGGE Project Office. The GARP Project at NASA will have the overall responsibility for the successful completion of the special effort.

In its original proposal that a special effort for editing and data enhancement be undertaken, the panel of ex-

perts on quality control went on record stating that ". . . this effort is *crucial* for achieving full impact of nonconventional observing systems." In addition to greatly improving the value of the space-based data, a significant added benefit will be that in the process of equipping and training groups in both NASA and NOAA to accomplish this task, an enormous amount of highly useful technological transfer will occur from the research community to the key user groups. The special effort represents an opportunity to take a significant step forward in improving the usefulness of satellite data in numerical weather prediction. For the first time the concepts of enhanced resolution in areas of greater meteorological interest, man/machine interactive processing, and internal data consistency (data source to data source) will be simultaneously brought into play. •

II. MODELING AND SIMULATION FACILITY

Paper 3

NASA Conference Publication 2076 – Fourth NASA Weather and Climate Program Science Review, January 24-25, 1979, NASA/GSFC, Greenbelt, MD.

Paper No. 25

CASE STUDIES OF MAJOR DST-6 SOUNDING IMPACTS WITH THE GLAS MODEL

R. Atlas, *Goddard Space Flight Center, Greenbelt, Maryland*

ABSTRACT

Two case studies of DST-6 sounding impacts are presented. In each of these cases, major improvements to the GLAS model's forecasts of specific synoptic features resulted from including satellite-sounding data in the initial analysis.

Paper 4

NASA Technical Memorandum 80591 November 1979

A Comparison of GLAS SAT and NMC High Resolution NOSAT Forecasts from 19 and 11 February 1976

Robert Atlas

Abstract This report presents the results of a subjective comparison of GLAS and NMC high resolution model forecasts. Two cases where NMC's operational model in 1976, had serious difficulties in forecasting for the United States have been examined: the 72 hour forecasts from 0000 GMT 19 and 11 February 1976. For each of these cases the GLAS model forecasts from initial conditions which included satellite sounding data were compared directly to the NMC higher resolution model forecasts, from initial conditions which excluded the satellite data. The comparison showed that the GLAS SAT forecasts significantly improved upon the current NMC operational model's predictions in both cases.

Paper 5

Reprinted from the Fourth Conference on Numerical Weather Prediction, October 29 – November 1, 1979, Silver Spring, MD.

SUBJECTIVE EVALUATION OF THE COMBINED INFLUENCE OF SATELLITE TEMPERATURE SOUNDING DATA AND INCREASED MODEL RESOLUTION ON NUMERICAL WEATHER FORECASTING

R. Atlas, M. Halem, M. Ghil*

NASA/Goddard Space Flight Center
Greenbelt, Maryland

* and Courant Institute of Mathematical Sciences
New York, New York

1. INTRODUCTION

Since the introduction of satellite derived temperature soundings into the operational data base in 1972, a number of experiments have been conducted [Atkins and Jones, 1975; Desmarais et al., 1978; Druyan et al., 1978; Ghil et al., 1979a, b] to assess their influence on numerical weather forecasting. Objective measures of forecast accuracy have served as the primary tool for evaluating the predictive impact of the satellite sounding data experiments, along with some subjective evaluations by weather forecasters. The statistical measures have generally shown either negligible or small beneficial impacts when averaged over many cases and continental regions. Even when specific forecasts show substantial statistical impacts, subsequent subjective evaluations have shown the improvements to be of marginal value for forecasters. A principal reason may have been the fact that these statistical improvements occur in the 48 to 72 hour forecast range when the model forecast accuracies have themselves deteriorated.

Thus, the question arises whether satellite sounding data might have greater influence, beneficial or otherwise, if NWP models had greater forecast skill. Recent studies by Robert [1976], Williamson [1978], Quirk and Atlas [1977] have shown that forecast accuracy is generally improved with increased model resolution. Our study here will report on the subjective comparisons of temperature sounding data impacts for models with two different grid resolutions, a coarse (400 km^2) and finer (280 km^2) resolution.

Atkins and Jones [1975] made a detailed subjective evaluation of the impact of Satellite Infrared Spectrometer (SIRS) sounding data on the numerical prediction of sea-level pressure and 500 mb heights for the period from 9-15 March 1974. Their evaluation revealed that no significant impacts occurred, however, slight improvements due to the inclusion of SIRS data were noted. Their assessment was performed by comparing prognostic charts generated from initial conditions which included satellite sounding data (referred to as SAT) with those generated from initial conditions which excluded satellite sounding data (referred to as NOSAT), and with the corresponding analyses for verification. Cases of significant impact were defined by Atkins and Jones as those in which different local weather forecasts would have resulted from use of the SAT prognoses.

In the more recent Data Systems Tests (DST) conducted by the U.S. National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA) in preparation for the Global Weather Experiment, subjective comparisons of SAT and NOSAT prognostic charts were made as part of the total evaluation of sounding data impact. Ghil et al. [1979a] utilized the Goddard Laboratory for Atmospheric Sciences (GLAS) general circulation model [Somerville et al., 1974; Stone et al., 1974] with a horizontal resolution of 4° latitude by 5° longitude to test different methods for the time continuous assimilation of the High Resolution Infrared Spectrometer (HIRS) and Scanning Microwave Spectrometer (SCAMS) data from Nimbus 6 and Vertical Temperature Profile Radiometer (VTPR) data from NOAA 4. Their evaluation for eleven SAT and NOSAT forecasts, generated during the February 1976 DST period showed that: (i) satellite-derived temperature data had a modest, but statistically significant positive impact on numerical weather forecasts, as verified over the continents of the Northern Hemisphere; (ii) the impact of a two-satellite system was about twice as large as that of one satellite; (iii) the magnitude of the impact depended upon the method of satellite data assimilation.

Desmarais et al. [1978], using the National Meteorological Center (NMC) six level primitive equation model [Shuman and Hovermale, 1968], found a smaller beneficial influence of satellite sounding data than what was obtained by Ghil et al., [1979a]. They argued that the difference between the GLAS and NMC results might be due to the poorer forecast skill of the GLAS model and analysis

system, thereby allowing for the satellite data to have a greater positive impact on the forecast.

A repetition of the forecast experiments for the February 1976 DST period, using the GLAS model with an increased horizontal resolution of 2.5° latitude by 3° longitude [Ghil et al., 1979b] have shown a substantial improvement in the forecast skill of the model. Objective measures of forecast accuracy of coarse and high resolution SAT and NOSAT prognoses indicated that both the increased resolution and the sounding data assimilation improved the forecast accuracy.

In this paper we present our subjective evaluation of the influence of satellite sounding data on the high resolution GLAS model forecasts. This evaluation was concerned with (1) the significance of prognostic differences resulting from the inclusion of satellite derived temperature soundings, (2) whether the differences were beneficial, (3) the dynamics of forecasting impact, i.e., how specific differences between the SAT and NOSAT prognoses evolved, and (4) comparison of this experiment with those previously reported for the 4° latitude by 5° longitude version of the GLAS model.

2. SUBJECTIVE EVALUATION

The subjective evaluation was performed by comparing the prognostic charts generated from SAT and NOSAT initial condictions with each other and with the corresponding NMC analyses for verification. This consisted of subjectively analyzing a variety of prognostic fields every 12 hours, for eleven different forecasts from the SAT and NOSAT cycles from the February 1976 DST period.

Tables 1 and 2 present the consensus of three experienced forecasters in their subjective assessment of the prognostic charts. During the evaluation, the forecasters had no knowledge of which prognostic charts were made from SAT or NOSAT initial conditions. Following Atkins and Jones [1975], each forecast was classified according to the following scale:

A - SAT significantly better than NOSAT,
B - SAT better than NOSAT,
C - SAT and NOSAT of equal quality,
D - SAT worse than NOSAT,
E - SAT significantly worse than NOSAT.

Each forecaster performed his evaluation independently. This was followed by a discussion of the prognostic differences, after which the ratings were averaged arrive at the consensus.

The results of our comparisons for the sea-level pressure and 500 mb height fields generated from the GLAS 4° latitude by 5° longitude model and verified over North America are presented in Table 1 [reproduced from Ghil et al. 1979a].

TABLE 1

RESULTS OF A SUBJECTIVE ASSESSMENT OF SAT AND NOSAT FORECASTS USING THE GLAS 4° BY 5° MODEL

a. SEA-LEVEL PRESSURE

Category	12 h	24 h	36 h	48 h	60 h	72 h	Total
A	0	0	0	0	0	0	
B	0	1	1	4	3	4	13
C	8	8	7	3	6	5	36
D	3	2	3	3	2	2	15
E	0	0	0	0	0	0	0

b. 500 MB GEOPOTENTIAL HEIGHT

Category	12 h	24 h	36 h	48 h	60 h	72 h	Total
A	0	0	0	0	0	0	0
B	1	1	2	3	4	4	15
C	8	9	9	5	7	6	44
D	1	0	0	3	0	1	5
E	0	0	0	0	0	0	0

TABLE 2

RESULTS OF A SUBJECTIVE ASSESSMENT OF SAT AND NOSAT FORECASTS USING THE GLAS 2.5° BY 3° MODEL

a. SEA-LEVEL PRESSURE

Category	12 h	24 h	36 h	48 h	60 h	72 h	Total
A	0	0	0	1	2	2	5
B	0	2	2	2	1	3	10
C	11	8	7	7	5	5	43
D	0	1	2	1	3	1	8
E	0	0	0	0	0	0	0

b. 500 MB GEOPOTENTIAL HEIGHT

Category	12 h	24 h	36 h	48 h	60 h	72 h	Total
A	0	0	0	1	1	2	4
B	0	1	3	2	3	3	12
C	11	9	7	7	6	7	47
D	0	1	1	1	1		4
E	0	0	0	0	0		0

From this table it can be seen that at sea level, none of the prognostic charts have been classified as A or E and that an almost equal number of B and D marks have been assigned. Although no significant impact has occurred at this atmospheric level, a tendency for more beneficial impacts at 48, 60, and 72 hours is evident.

At 500 mb, two 72-hour prognostic charts have been classified as A and there were no E marks. In addition more than three times as many B than D marks have been assigned. Impacts at this level are most consistently beneficial at 60 and 72 hours.

Table 2 presents the results of our subjective evaluation of sounding data impact using the GLAS 2.5° by 3° model. This table shows that a more substantial beneficial impact from satellite sounding data has resulted from the use of the higher resolution model. No significant negative impacts have occurred while five sea-level pressure and four 500 mb prognostic charts have been significantly improved. Once again the effect of satellite data is largest during the latter half of the forecast period.

3. EXAMPLES OF FORECAST IMPACT

Two cases of significant forecast impact, the forecasts from 19 and 11 February 1976, were selected for a more detailed subjective analysis. The 72-hour sea-level pressure prognoses from these cases are presented here as illustrations of the combined influence of satellite data and increased resolution.

a. 72-Hour Forecast from 0000 GMT, February 19, 1976

This was a case in which a significant improvement to the predicted displacement of an intense winter storm occurred in the latter part of the forecast period. The initial conditions for this forecast showed a moderately intense low pressure system, associated with an upper level short wave trough, located off the northwest coast of the U. S. As the system moved inland, a new low developed along an already existing stationary front, and became the dominant feature by 1200 GMT on February 19. During the next 24 hours, the low moved southeastward and intensified, after which time it recurved and then accelerated toward the northeast. This cyclone was accompanied by heavy snow, blizzard, or near-blizzard conditions in Colorado, Kansas, Nebraska, Iowa, Michigan, and Wisconsin, and tornadoes or severe thunderstorms in Kansas, Oklahoma, Texas, Arkansas, Missouri, Illinois, Louisiana, and Mississippi. Most of the severe weather occurred toward the end of the forecast period after the recurvature of the cyclone to the northeast.

Figs. 1a and 1b depict the 72-hour SAT and NOSAT sea-level pressure prognoses for the 4° by 5° model, while the corresponding SAT and NOSAT prognoses for the 2.5° by 3° model are shown in Figs. 1c and 1d. The verifying analysis is shown in Fig. 1e. Comparison of these charts reveals little difference between the two 4° by 5° model forecasts, however a very significant improvement in the predicted displacement of the surface low has resulted from the inclusion of satellite data in the 2.5° by 3° model. The high resolution SAT prediction indicated recurvature of the low to the northeast while the other predictions did not. As a result, there was a more than 1000 km reduction in the vector error of the predicted position.

Since most of the severe weather occured in conjunction with the recurvature of the low, it is clear that significantly improved local weather forecasts could have resulted from the use of the high resolution SAT prognoses in this situation. To investigate this point further, the high resolution SAT and NOSAT predictions of local convective instability, and also the destabilization by differential equivalent potential temperature advection were compared. It was found that there was a 51 percent reduction in the RMS errors of convective instability and a 22 percent reduction in RMS errors of differential advection, in the prediction which included satellite-sounding data.

b. 72-Hour Forecast from 0000 GMT, February 11, 1976

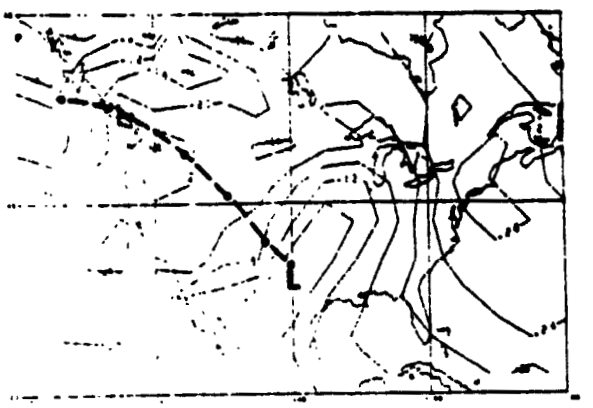
In this case, a weak cyclone formed along a stationary front in southwest Canada, moved southeastward while intensifying during the first 40 hours of the period, and then recurved to the east northeast.

Figs. 2a and 2b depict the 72-hour SAT and NOSAT sea-level pressure prognoses for the 4° by 5° model, while the corresponding SAT and NOSAT prognoses for the 2.5° by 3° model are presented in Figs. 2c and 2d. Fig. 2e shows the verifying analysis. Comparison of these figures reveals that an improvement in the predicted displacement of the surface low has resulted from the inclusion of satellite data in the 4° by 5° model, but a much larger improvement has occurred in the 2.5° by 3° model. The high resolution SAT prognosis (Fig. 2c) is substantially better than the NOSAT (Fig. 2d) in its prediction of the cyclonic circulation over southeastern Canada and the northeastern U.S., as well as the anticyclone near the Great Lakes and the pressure trough which extends southeastward from southwest Canada.

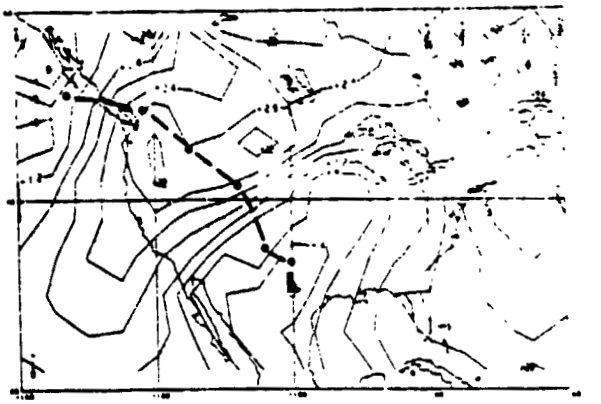
4. EVOLUTION OF FORECAST IMPACT

In this section the development and growth of the forecast differences between the GLAS high resolution SAT and NOSAT forecasts from February 19 will be described. A similar study for other model forecasts and for the February 11 case is currently being prepared.

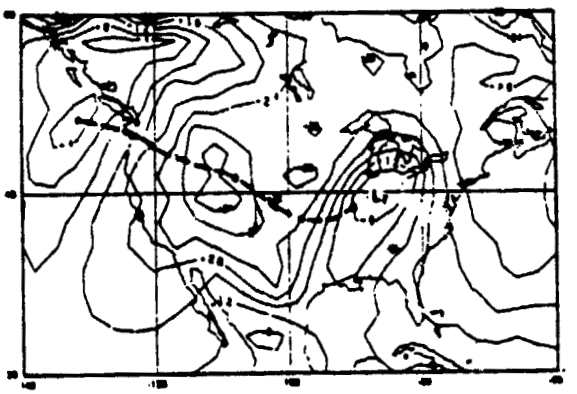
Comparison of Figs. 1c and 1d reveals that the predicted movements of the surface low are similar for the first thirty-six to forty-eight hours of the forecast but then diverge rapidly. We are particularly concerned with explaining (1) why the surface lows in the two forecasts suddenly diverge and move differently during the latter half of the forecast and (2) if differences in the forcing for the cyclone at the time the two forecasts diverge, can be traced to specific initial state differences between the SAT and NOSAT systems.



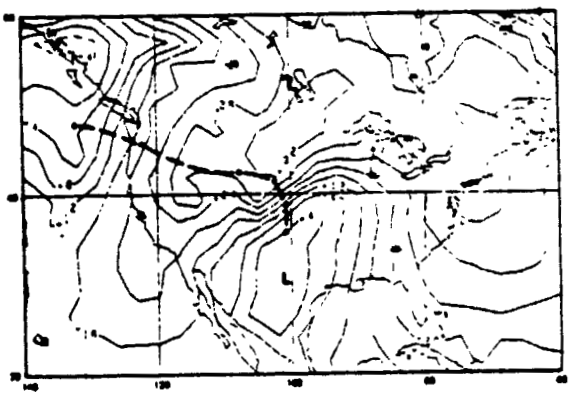
a



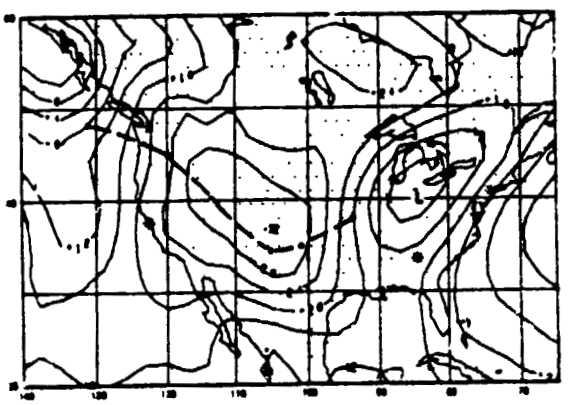
b



c



d



e

Fig. 1 The sea level pressure fields on 0000 GMT 22 February 1976

a. the 72 hour forecast from the 4° by 5° GLAS model with SAT initial conditions;

b. the corresponding NOSAT forecast using the 4° by 5° model;

c. the corresponding SAT forecast using the 2.5° by 3° GLAS model;

d. the corresponding NOSAT forecast using the 2.5° by 3° GLAS model;

e. the verifying NMC analysis. Dots represent past positions of the cyclone center at 12 hour intervals beginning 0000 GMT 19 February 1976.

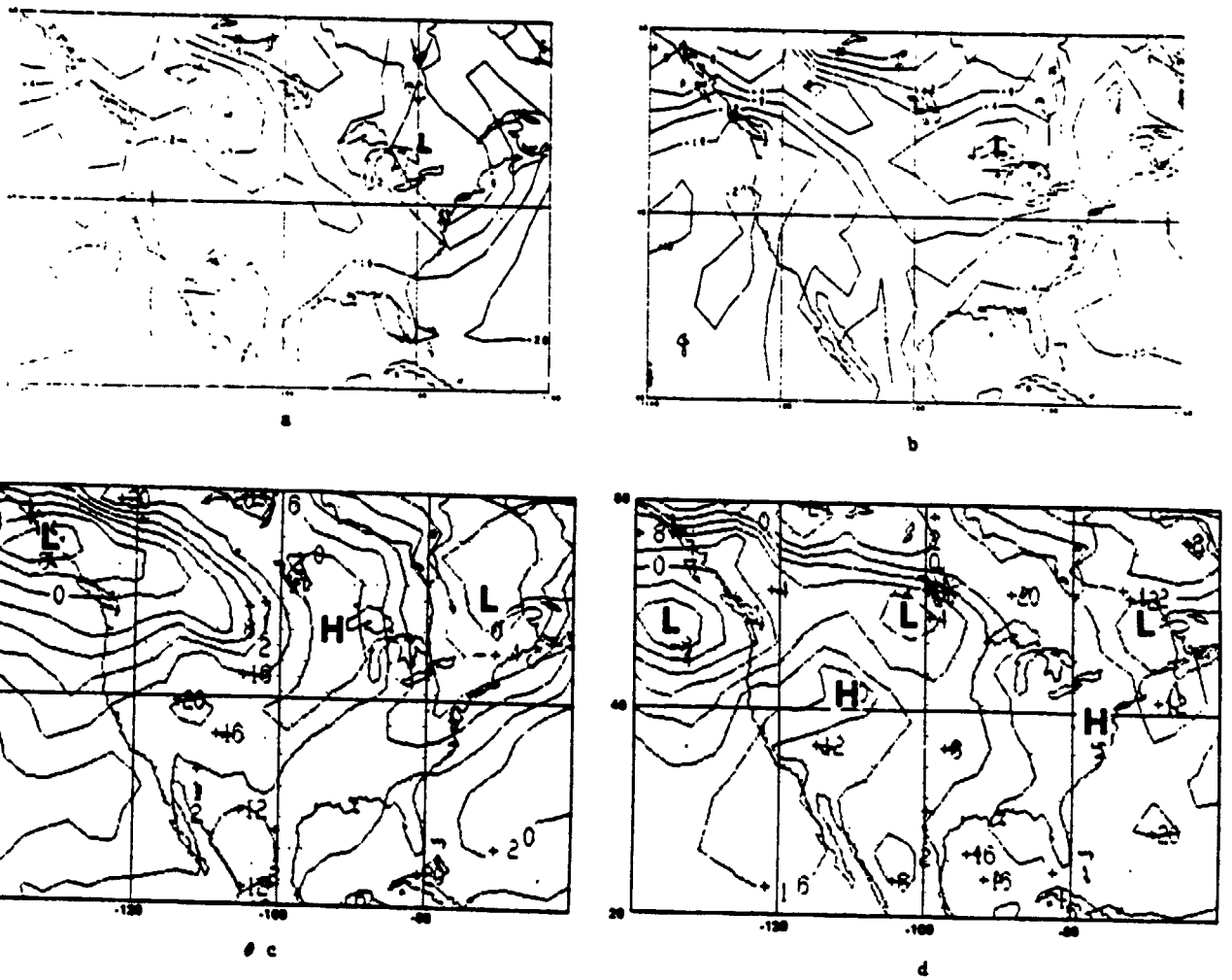


Fig. 2 The sea level pressure fields on 0000 GMT 14 February 1976

a. the 72-hour forecast from the 4° by 5° model with SAT initial conditions;

b. the corresponding NOSAT forecast using the 4° by 5° GLAS model;

c. the corresponding SAT forecast using the 2.5° by 3° GLAS model;

d. the corresponding NOSAT forecast using the 2.5° by 3° GLAS model;

e. the verifying NMC analysis.

To answer these questions we have examined horizontal distributions and vertical cross sections of a wide range of primary variables and model derived quantities related to the dynamics of the cyclone. Only a few of the more descriptive aspects of this study are presented here.

The initial sea level pressure and 1000-500 mb thickness, and 300 mb height and vorticity analyses, for the SAT and NOSAT forecasts from February 19, are displayed in Figs. 3 and 4. Comparison of these figures reveals no substantial difference in the initial representation of the surface low located at 46°N, 132°W off the northwest coast of the U. S., although a slight modification of the 1000-500 mb thickness pattern has occurred. This results in a slightly enhanced variation of thermal vorticity and thermal advection across the low center in the SAT case. At 300 mb larger differences between the two initial states are evident.

The assimilation of satellite sounding data has resulted in an intensification of the upper level vorticity maximum associated with the surface low and an elimination of the vorticity trough extending southward from this maximum. The satellite data has increased the northerly component of the wind and the cyclonic shear of the westerly component of the wind from 40°N to 47°N to the immediate west of the upper level trough. It also increased the anticyclonic shear of the westerly wind below 40°N (wind component charts not shown). This results in a northward shift of the upper level vorticity advection areas in this region, such that there is stronger positive vorticity advection and positive vorticity tendencies to the east of the 300 mb vorticity maximum, and stronger negative vorticity advection and negative vorticity tendencies to the west of the maximum in the SAT case. This modification of the vorticity advection by satellite sounding data was found to be in agreement with satellite cloud imagery, (see Hales 1979 for discussion of technique). Additional modifications to the initial state by the assimilation of satellite data, have been considered in this study but will not be described here.

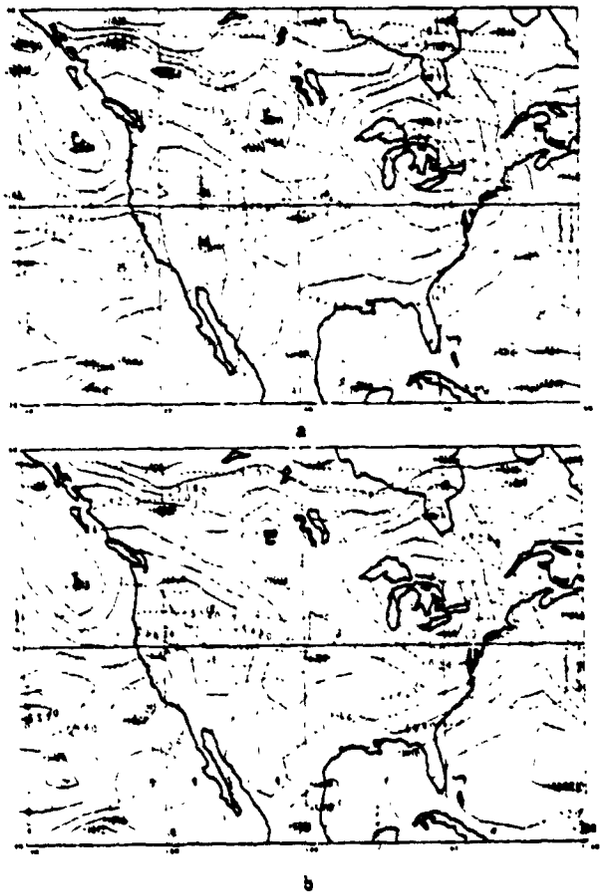


Fig. 3 SAT (a) and NOSAT (b) Sea-level pressure (solid lines) and 1000-500 mb thickness (dashed lines) for 0000 GMT 19 February 1976.

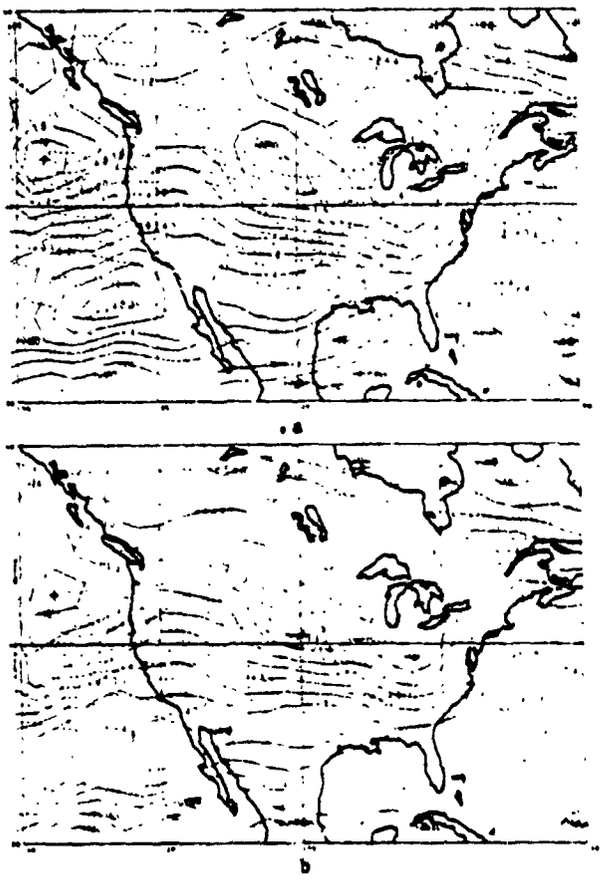


Fig. 4 SAT (a) and NOSAT (b) 300 mb absolute vorticity (solid lines) and 300 mb geopotential height (dashed lines) for 0000 GMT 19 February 1976.

Figs. 5-12, display the evolution of the sea-level pressure, 1000-500 mb thickness, and 300 mb height and vorticity patterns at twelve hour intervals for the first 48 hours of both the SAT and NOSAT predictions. During the first 12 hours of the forecast both SAT and NOSAT move the surface low inland to 46°N, 123°W (Fig. 5). The SAT surface low has intensified while the NOSAT low has filled slightly. However the SAT thickness gradient has weakened relative to the NOSAT and there is now a larger variation of thermal vorticity and stronger thermal vorticity advection across the low in the NOSAT case. At 300 mb a small difference in the movement of the height trough and vorticity maximum has occurred (Fig. 6). The NOSAT vorticity maximum is located at 44°N, 128°W while the SAT vorticity maximum has moved about 2.5° longitude further downstream. A comparison of the vorticity advection patterns and tendencies at two hour intervals throughout this period indicated that the initial tendency for greater movement in the SAT system has been maintained.

During the next twenty-four hours of the forecast this trend continues and slightly larger differences between the SAT and NOSAT systems begin to develop. At 0000 GMT 20 February (Figs. 7 and 8), the NOSAT system has forecast the surface low to deepen slightly and move to 41.3°N, 108°W, and the 300 mb vorticity maximum to move to 42°N, 123°W. SAT forecasts a less organized surface low with three centers evident at this time, but has maintained a somewhat stronger pressure gradient than the NOSAT to the west and southwest of the southernmost low center. SAT forecasts the 300 mb vorticity maximum to move to 41°N, 117°W, 5° longitude further downstream than the NOSAT maximum. However in both forecasts the strongest positive vorticity advection is to the southwest of the surface low position. By 1200 GMT 20 February the NOSAT surface low is located at 41.2°N, 105°W while SAT forecasts two low centers: one at 41.2°N, 102°W; the other at 36.2°N, 102°W (Fig. 9). Substantial differences in the low level thermal advection patterns are also evident at this time. In the NOSAT case, weak cold advection exists directly behind the surface low, with the strongest cold advection located further southwest beneath and slightly to the west of the upper level trough. This contributes to the deepening and slow rate of movement of the upper level trough in the NOSAT prediction. SAT predicts substantially stronger cold advection directly behind the surface low.

The relationship of low level thermal advection to the movement of surface lows is well known [Sutcliffe, 1947; Petterssen, 1954]. Warm advection contributes to falling pressure and the generation of cyclonic vorticity in advance of cyclones while cold advection contributes to rising pressure and the destruction of cyclonic vorticity to the

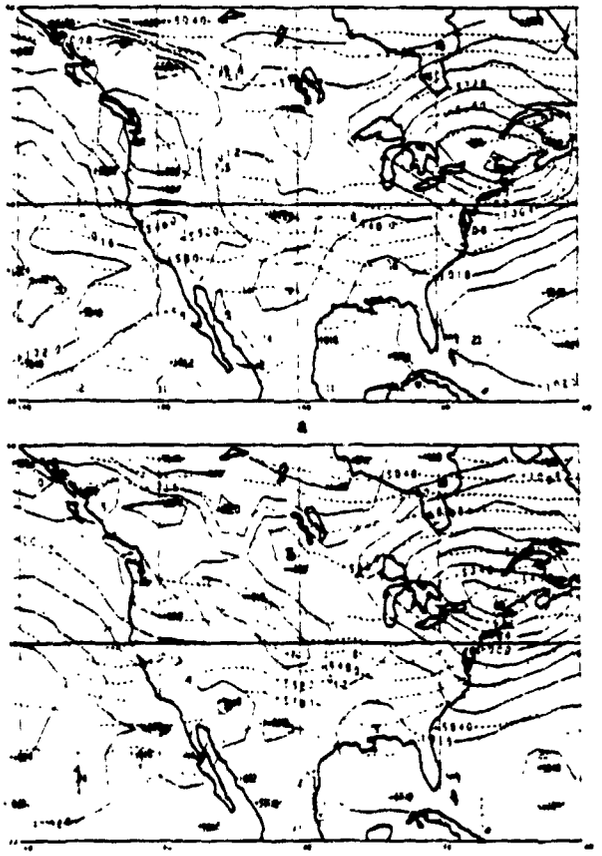


Fig. 5 Same as Fig. 3 except for 1200 GMT 19 February 1976.

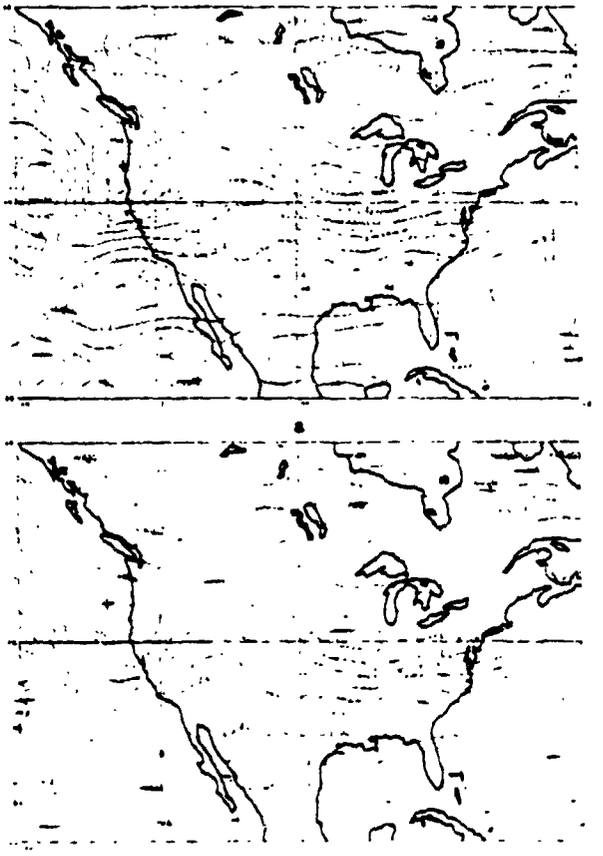


Fig. 6 Same as Fig. 4 except for 1200 GMT 19 February 1976.

rear of cyclones. Thus the gradient of thickness advection across a cyclone center is an important indicator of the rate of movement of that cyclone. In this case, the differences in the gradient of thickness advection that are established after thirty-six hours are primarily associated with the differing rates of movement of the upper level vorticity maximum. Because of the different movements of the vorticity maximum relative to the surface low, there was an intensification of the pressure gradient to the southwest of the low center in the SAT forecast, while a broad area of surface pressure falls associated with the positive vorticity advection aloft weakened the pressure gradient to the immediate southwest of the surface low in the NOSAT forecast.

This effect amplified during the next twelve hours in such a way that by 0000 GMT 21 February the gradient of thickness advection across the low center is nearly twice as strong in the SAT forecast as in the NOSAT. At this time the surface low is located at 36.5°N, 102°W in the NOSAT and 36°N, 99°W in the SAT (Fig. 11). The 300 mb vorticity maximum is located at 36.2°W, 111°W in the NOSAT and 38°N, 102°W in the SAT (Fig. 12). The differing phase relationships between the upper level vorticity maximum and the surface low coupled with the differing thickness advection patterns result in the diverging paths of the surface low throughout the remainder of the forecast. In the SAT case, there is strong positive vorticity advection and warm advection to the east and northeast of the surface low and weaker vorticity advection coupled with strong cold advection to the immediate west and southwest of the surface low. This results in a strong isallobaric gradient across the low center such that recurvature to the east northeast occurs. In the NOSAT case, weak positive vorticity advection is coupled with warm advection to the northeast of the surface low while strong positive vorticity advection is coupled with weak cold advection to the immediate west and southwest of the surface low. As a result the NOSAT surface low becomes "locked in" and does not progress eastward after this time.

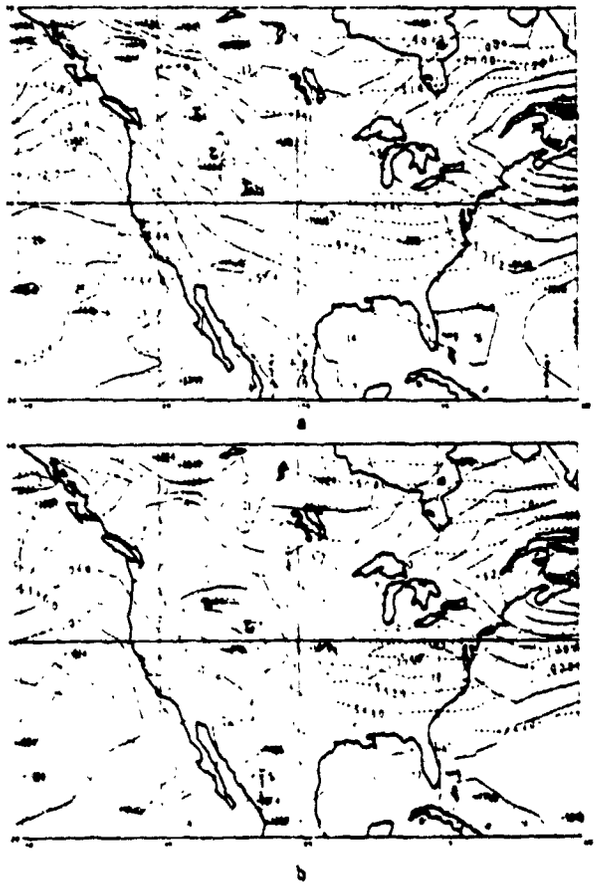


Fig. 7 Same as Fig. 3 except for 0000 GMT 20 February 1976.

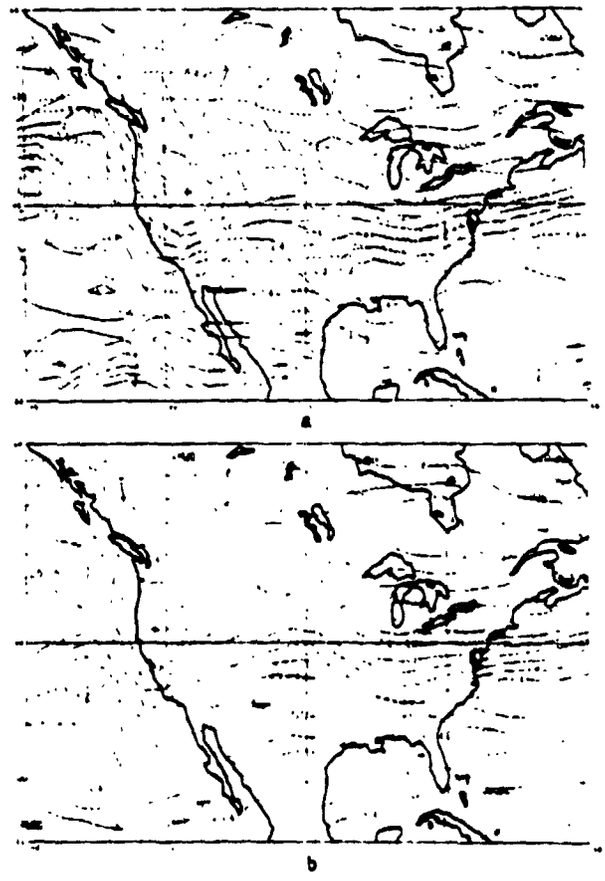


Fig. 8 Same as Fig. 4 except for 0000 GMT 20 February 1976.

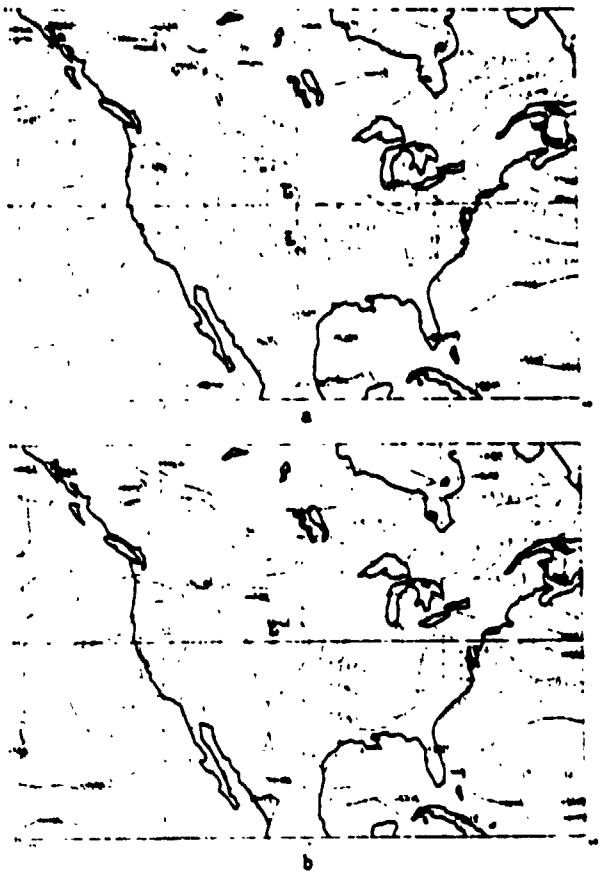


Fig. 9 Same as Fig. 3 except for 1200 GMT 20 February 1976.

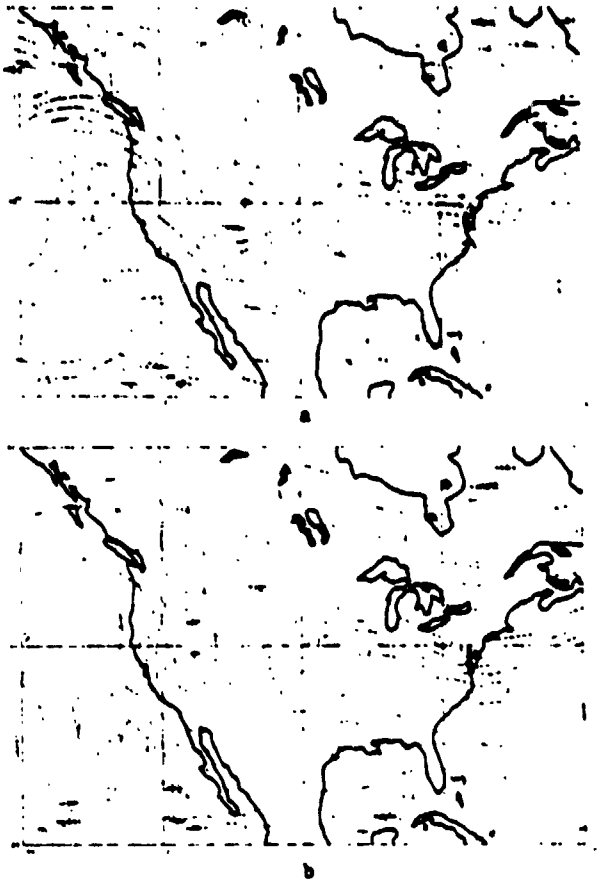


Fig. 10 Same as Fig. 4 except for 1200 GMT 20 February 1976.

5. CONCLUSIONS

The subjective evaluation of the influence of satellite sounding data on forecasts with the GLAS model has shown that the modest beneficial impact of sounding data has been enhanced by the increased resolution. Two cases, where significant differences occur in the high resolution model forecasts, were described.

In one of these cases, the forecast impact was traced to an initial state modification by the sounding data which was consistent with satellite cloud imagery. The other case study analysis is as yet incomplete.

While questions of the model dependency of these forecast impacts have not been fully explored, this study suggests that satellite sounding data possess valuable information content which at times are clearly capable of correcting gross analysis errors in data sparse regions.

ACKNOWLEDGMENT

The authors wish to acknowledge R. Rosenberg and M. Eaton for their assistance in the evaluation of sounding data impact, and D. Sakal and S. Palm for providing computer and graphics support.

REFERENCES

Atkins, M. J., and M. V. Jones, 1975: An experiment to determine the value of satellite infrared spectrometer (SIRS) data in numerical forecasting. Meteor. Mag., 104, 125-142.

Desmarais, A., S. Tracton, R. McPherson, and R. Van Haaren, 1978: The NMC report on the Data Systems Test. NOAA-ST 78-218, National Meteorological Center, NOAA/ National Weather Service, 331 pp.

Druyan, L. M., T. Ben-Amram Z., Alperson, and G. Ohring, 1978: The impact of VTPR data on numerical forecasts of the Israel Meteorological Service. Mon. Wea. Rev., 106, 859-869.

Ghil, M., M. Halem, and R. Atlas, 1979a: Time-continuous assimilation of remote-sounding data and its effect on weather forecasting. Mon. Wea. Rev., 107, 140-171.

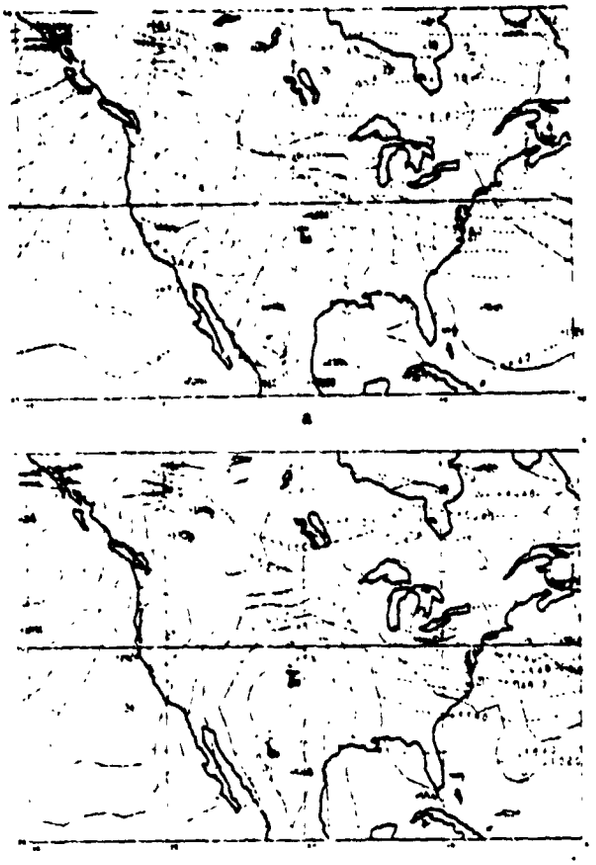


Fig. 11 Same as Fig. 3 except for 0000 GMT 21 February 1976.

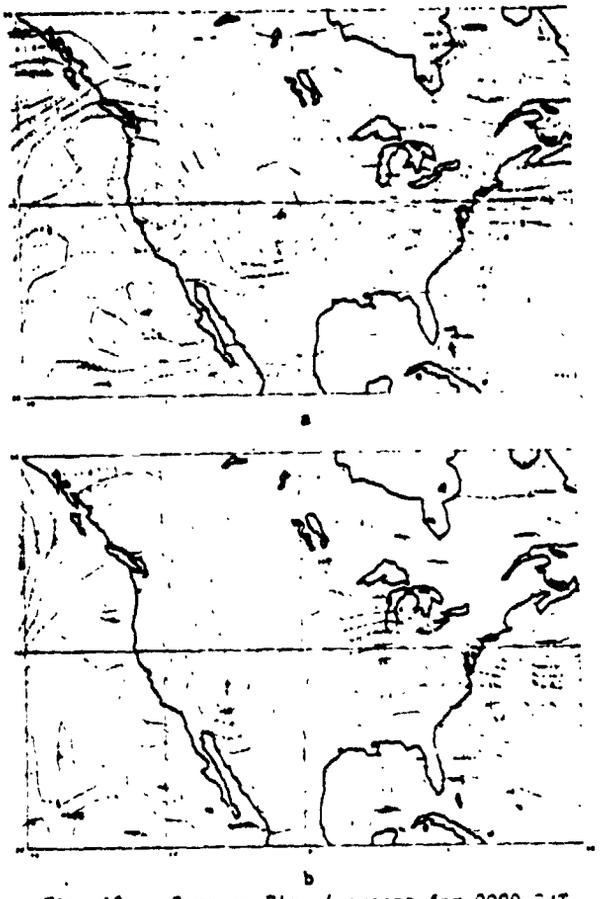


Fig. 12 Same as Fig. 4 except for 0000 GMT 21 February 1976.

_____, 1979b: Effects of sounding temperature assimilation on weather forecasting: model dependence studies. Remote Sounding of the Atmosphere from Space, H. J. Bolle, ed., Permagon Press, Oxford and New York, 21-25.

Halem, M. E., 1979: A subjective assessment of model initial conditions using satellite imagery. Bull. Amer. Meteor. Soc., 60, 206-211.

Petterssen, S., 1956: A general survey of the factors influencing development at sea level. J. Meteor., 12, 36-42.

Quirk, W. J. and R. Atlas, 1977: The effect of increased horizontal resolution on synoptic forecasts with the GISS model of the global atmosphere. Conf. Papers, Third Conference on Numerical Weather Prediction, 92-99.

Robert, A., 1976: Sensitivity experiments for the development of NWP models. Proc. of the (Eleventh) Stanstead Seminar, Publication in Meteorology No. 114, McGill University, Montreal, Canada, 68-81.

Somerville, R. C. J., P. H. Stone, M. Halem, J. E. Hansen, J. S. Hogan, L. M. Druyan, G. Russell, A. S. Lacis, W. J. Quirk and J. Tenenbaum, 1974: The GISS model of the global atmosphere. J. Atmos. Sci., 31, 84-117.

Stone, P. H., S. Chow and W. J. Quirk, 1977: The July climate and a comparison of the January and July climates simulated by the GISS general circulation model. Mon. Wea. Rev., 105, 170-194.

Sutcliffe, R. C. 1947: A contribution to the problem of development. Quart. J. R. Meteor. Soc., 73, 370-383.

Williamson, D. L., 1978: The relative importance of resolution, accuracy, and diffusion in short-range forecasts with the NCAR global circulation model. Mon. Wea. Rev., 106, 69-88.

NASA Conference Publication 2076 – Fourth NASA Weather and Climate Program Science Review, January 24-25, 1979, NASA/GSFC, Greenbelt, MD.

Paper No. 23

SENSITIVITY OF FORECAST SKILL TO DIFFERENT OBJECTIVE ANALYSIS SCHEMES

W. E. Baker, *National Research Council, Washington, DC and Goddard Space Flight Center, Greenbelt, Maryland*

ABSTRACT

Numerical weather forecasts are characterized by rapidly declining skill in the first 48 to 72 h. Recent estimates of the sources of forecast error indicate that the inaccurate specification of the initial conditions contributes substantially to this error.

The sensitivity of the forecast skill to the initial conditions is examined by comparing a set of real-data experiments whose initial data were obtained with two different analysis schemes. Results are presented to emphasize the importance of the objective analysis techniques used in the assimilation of observational data.

Paper 7

Reprint from Journal of Marine Research, Volume 37, 2, 1979.

The response of an equatorial ocean to simple wind stress patterns: I. Model formulation and analytic results

by Mark A. Cane[1]

ABSTRACT

A simple model is developed to study the wind-driven equatorial ocean circulation. It is a time dependent, primitive equation, beta plane model that is two-dimensional in the horizontal. The vertical structure consists of two layers above the thermocline with the same constant density. The ocean below the thermocline is taken to be of a higher constant density and to be approximately at rest. The surface layer is of constant depth and is acted upon directly by the wind. The depth of the lower active layer is dynamically determined. This is the simplest vertical structure which allows a steady state undercurrent.

The linear dynamics of the model are investigated analytically. The circulation evolves on two timescales: a frictional component is established in 0 (20 days) while the pressure gradients and interior flow have a longer, highly variable setup time. The steady transport consists of a Sverdrup part and a frictionally driven part in the vicinity of the equator. When forced with a uniform easterly wind, the flow at the equator in the lower layer is in the same direction as the undercurrent (eastward). However, the vertically integrated transport is westward. This differs from observations and suggests that inertial effects must be included to properly simulate the undercurrent.

In a companion paper (Cane, 1979) both the linear and nonlinear dynamics of the model are investigated by numerical methods.

1. Introduction

Since the vertical component of the Coriolis force vanishes at the equator, the geostrophic balances which dominate the dynamics of the extra-equatorial oceans must break down. The most striking physical manifestation of this singularity is the Equatorial Undercurrent, a narrow (half width of 1°), fast (speeds up to 170 cm/sec), eastward flowing subsurface current in the thermocline of all the world's oceans. (While it is a permanent feature in the Atlantic and Pacific at most longitudes, it has been observed only intermittently in the Indian Ocean, Knox, 1976.) Many of the characteristics of the undercurrent are highly variable: e.g., the downstream velocities and transports may vary by a factor of two or more at different longitudes or at different times. Available observational data allow many of these variations to be

1. NASA/Goddard Space Flight Center, Laboratory for Atmospheric Sciences, Greenbelt, Maryland, 20771, U.S.A.

related systematically to variations in the winds over the equatorial ocean. However, the evidence is, in general, too spotty to allow such correlations to be conclusive. (Philander (1973b) presents a thorough review of the measurements of the undercurrent made up to 1973.)

A second important consequence of the vanishing of the Coriolis term is that equatorial motions have time scales which are very much shorter than those of midlatitude motions: the baroclinic time scale is weeks at the equator, as against years at midlatitudes. The most impressive instance of this short time scale is the reversal in direction of the Somali Current within a month of the onset of the Southwest Monsoon (e.g., Leetmaa, 1973).

Because of the rapidity of the response to atmospheric forcings, equatorial oceans are rewarding areas for the study of motions with time scales of a month or longer. Until quite recently such motions received little theoretical attention. Much of the work in this area has followed Lighthill (1969) in focusing on the set-up of the Somali Current in response to the onset of the Southwest Monsoon. The time dependent behavior of the undercurrent itself has received far less attention. Gill (1975) applied a Lighthill-like model to the undercurrent in the western Pacific. He associated the undercurrent with the second baroclinic mode Kelvin wave which propagates in from the western boundary. It is not clear how such a model explains the presence of the undercurrent as a more permanent feature. Philander (1976) explained the undercurrent meanders observed during GATE (Düing *et al.*, 1975) as the result of a shear instability of the surface currents.

In contrast to the situation for time varying equatorial currents, numerous theoretical models for the steady state undercurrent appear in the literature. These have been reviewed by both Gill (1975) and Philander (1973b). For this reason we shall forego a detailed review here: rather, we shall discuss them only to the extent needed to establish a theoretical context for the present work. On the basis of his observations in the Pacific, Knauss (1966) estimated that the only negligible terms in the momentum equation were those giving the time rate of change of momentum and the horizontal component of the Coriolis force due to vertical motion. (He did not consider horizontal eddy diffusion processes.) The upshot is that a great variety of processes are available to be used as explanations for the undercurrent. Since there is a certain amount of freedom in the choice of eddy coefficients, all of these can be expected to give agreement with at least some of the observed scales. In what follows, we seek to isolate those processes which are most significant.

We shall immediately restrict ourselves to those models which idealize the thermocline as a discontinuity between a shallow upper homogeneous layer and a deeper lower homogeneous layer of greater density. The lower layer is assumed to be so deep that its horizontal pressure forces and velocities vanish. Models with thermohaline components (Robinson, 1960; Philander, 1971, 1973a) are required to explain certain effects at depth; for example, the double celled structure often observed

in the Pacific (see Philander, 1973b). Homogeneous models appear to be sufficient for explaining observed features above the thermocline.

The most basic physical notion about the undercurrent is the idea of flow down a pressure gradient (Charney, 1960). The prevailing easterly winds pile up water at the western side of the ocean basin, thus establishing an eastward pressure gradient. Stommel (1960) exploited this idea to obtain an eastward flowing subsurface current in a linear model with vertical friction. He assumed free slip boundary condition at the bottom and that the vertically integrated transport vanishes at the equator. In a similar model without the latter two assumptions, Charney (1960) and Philander (1971) found that the current at the equator did not reverse with depth. In any case, one would wish any theory to account for the substantial eastward transports observed at the equator. In the linear theory of Gill (1971), the pressure gradient force is balanced by the horizontal mixing of momentum. By using an unrealistically large value for the coefficient of horizontal eddy viscosity (10^8 cm^2 sec^{-1}), Gill obtains the observed latitudinal scale for the undercurrent, but the transport is too low by a factor of at least four.

Nonlinear theories have ignored the downstream inertial terms. The (suspect) assumption is made that the zonal and meridional velocities have the same scale. Then, since the meridional length scale (an equatorial boundary layer scale) is so much shorter than the zonal one (the length of the basin), it follows that in the momentum equation the downstream inertial term is negligible relative to the cross-stream inertial term. Attention is then directed to the meridional circulation. For an easterly wind, the Ekman drift in the surface layers will be poleward. Continuity then requires a compensatory equatorward mass flux at depth, producing an upwelling region at the equator to complete the fluid circuit. Fofonoff and Montgomery (1955) considered the subsurface flow in the light of the barotropic vorticity equation. If it is assumed that a parcel approximately conserves the vertical component of its absolute vorticity, it must change its relative vorticity to make up for the loss of planetary vorticity as it moves equatorward. This results in an eastward flow at the equator. It may also be shown that the meridional circulation near the equator enhances the eastward transport at the equator regardless of whether the wind is easterly or westerly. (See Robinson (1966) for an analytic demonstration; Gill (1975) gives a more physical argument.)

The models of Charney (1960), Charney and Spiegel (1971), Robinson (1966), and McKee (1973) all incorporate the nonlinear effects due to the circulation in the meridional plane. The first three include momentum mixing in only the vertical direction. McKee's model is an extension of Gill's (1971) model into the nonlinear regime; horizontal eddy viscosity is the important frictional force here. A more realistic value for the zonal velocity is obtained compared to the linear model, but an unreasonably large value for the eddy coefficient is again used (10^8 cm^2 sec^{-1}) to obtain the observed undercurrent width. The models of Charney (1960) and

Charney and Spiegel (1971) (the first calculates the flow only at the equator by assuming it is an axis of symmetry; the second paper extends the first model to a meridional plane) give the observed undercurrent velocity and width using a value for the vertical eddy viscosity coefficient (15 $cm^2\ sec^{-1}$) in agreement with existing observational evidence (See Sec. 2). This model also gives good agreement with the observed vertical profile of the undercurrent. Vertical viscosity must be of some importance at depth in order to obtain a nonconstant profile below the boundary layer. Most importantly, a mechanism for the vertical exchange of momentum is needed to introduce the wind stress into the water. There is no similar logical necessity for introducing a significant amount of horizontal mixing. Further, there is no evidence that modeling such mixing gives better agreement with observations.

Previous work thus shows that it is necessary to consider vertical eddy viscosity and inertial effects but not lateral eddy viscosity in order to model the undercurrent effectively. As noted above, all of these models neglect any variation in the zonal direction (except that the zonal pressure gradient is taken as constant). This makes it impossible to ask a number of interesting questions; for example, one cannot investigate the undercurrent meanders observed during GATE. More generally, the issue of the relation of the undercurrent to the entire equatorial current system cannot be explored without considering the whole ocean basin. Since there is a substantial eastward transport at the equator, there must be compensating westward flow elsewhere in the ocean basin. Further, many time varying effects are inseparable from zonal variations. For example, the length of time it takes for the sea surface to set up from rest in response to a wind stress is determined by the speed of waves which propagate in from the boundaries of the basin.

In order to investigate questions of this sort, our model will be time dependent and two dimensional in the horizontal. Since the phenomena of interest are confined to an area near the equator, the basin need not have a great latitudinal extent; 15S to 15N has proven to be sufficient. The model equations are solved numerically because it is imperative that they be fully nonlinear.

In order to make it practical to perform many numerical integrations, the vertical structure is drastically simplified. It consists of two layers above the thermocline with the same constant density. The ocean below the thermocline is taken to be of a higher constant density and to be approximately at rest. The upper of the two active layers is a constant depth surface layer which is acted upon directly by the wind stress. The lower active layer is not directly affected by the wind. Its depth is variable, with the variations being dynamically determined. The two layers communicate via the vertical velocity at their interface as well as being frictionally coupled. *This is the simplest vertical structure that allows a steady undercurrent.*

Of course, this simplification prevents the simulation of the detailed vertical structure of the undercurrent. It is not our intention to do such numerical simulations. Previous work (especially Charney and Spiegel, 1971) provides a bridge for relating

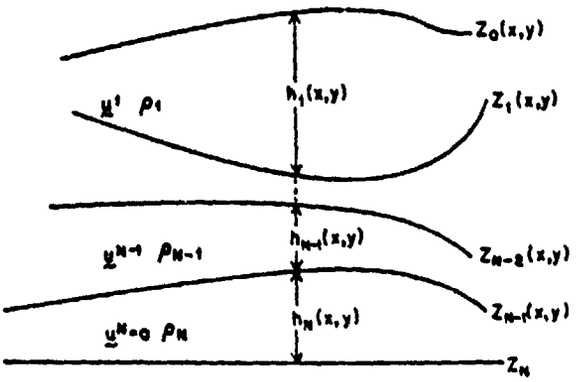


Figure 1. Multi-layer model.

the results of our simple model to the real world. Our philosophy is to treat the numerical experiments in the manner of laboratory experiments: we do not seek to simulate the real world; we seek merely to preserve enough analogy to the real world for the results to give insight into natural phenomena.

There are a large number of phenomena which may be investigated with such a model. In the present study we impose very simple wind stress patterns and study the evolution from a state of rest and eventual steady state configuration of the model ocean. In this paper the model is formulated and its linearized dynamics are explored by analytic means. Numerical results are discussed in the sequel to this work, Cane (1979), henceforth referred to as II. The analytic results are of interest in their own right in addition to being an invaluable aid in the interpretation of the numerical experiments. The vertical structure of the model is novel and these results help to clarify the relationship of the model variables to more familiar oceanographic models. They provide a descriptive vocabulary, a check on the numerical results, and a contrast that highlights the nonlinear effects.

2. Formulation of the physical model

The equations and parameters governing the model are discussed in this section. We begin by considering the familiar layered model in order to see why it is inappropriate for modeling the steady undercurrent. The equations for the model structure employed here are then derived, after which the choice of parameter values is discussed.

a. Layered model. Since we are concerned with inertial and viscous dynamics of a wind-driven ocean, thermohaline effects will be ignored. We divide the ocean vertically into stable material layers of constant density which are assumed to be nonmixing (Fig. 1). We now identify the bottom layer with the water mass below the thermocline and regard it as being sufficiently deep so that its velocity vanishes.

Equatorial regions are a favorable environment for this approximation: the thermocline is shallow (150-200 m), the wind stress projects about twenty times more strongly on the first baroclinic mode than it does on the barotropic mode (Lighthill, 1969), and unlike midlatitudes (Veronis and Stommel, 1956), the baroclinic signals are only about one order of magnitude slower than the barotropic. Further theoretical support may be drawn from Philander's (1977) analysis of vertical wave propagation, which shows that for the time scales of interest to us, most of the energy put in by the wind will remain above the thermocline. Observational evidence also tends to support the validity of this approximation (see Philander, 1973b for a summary).

Since the velocities in the lowest layer are assumed to vanish, the pressure gradient must vanish there as well. This leaves us with the familiar "reduced gravity" model; for a single active layer the equations governing the *average* horizontal current and the layer depth h_1 are

$$\frac{\partial}{\partial t}\,\mathbf{u}^1 + (\mathbf{u}^1 \cdot \nabla)\,\mathbf{u}^1 + f\,\mathbf{k} \times \mathbf{u}^1 = -g'\,\nabla h_1 + (\tau_0 - \tau_1)\,h_1^{-1} + \nu_H\,\nabla^2\,\mathbf{u}^1 \tag{2.1}$$

$$\frac{\partial}{\partial t}\,h_1 + \nabla \cdot (h_1\,\mathbf{u}^1) = 0. \tag{2.2}$$

Where the reduced gravity $g' = g\,(\rho_2 - \rho_1)/\rho_2$, τ_0 is the wind stress, τ_1 the interfacial stress between the two layers and $\nu_H \nabla^2 \mathbf{u}^1$ is the horizontal eddy stress.[2] The components of the Coriolis force due to vertical motions and departures from hydrostatic balance are neglected; this may be justified a posteriori.

The wind stress appears as a body force in (2.1). This is a commonly used modeling procedure in oceanography that can be rigorously justified for many purposes (e.g. Charney, 1955). However for some purposes, such as modeling the undercurrent, a difficulty is created by introducing the wind stress as a body force averaged over the uppermost layer.

For a constant easterly wind stress (of magnitude τ per unit mass) the steady state solution to (2.1), (2.2) is

$$\mathbf{u}^1 = 0;\; h_1^2 = h_1^2\,(x{=}0) + x\,\tau/g'. \tag{2.3}$$

The zonal pressure gradient is balanced by the wind stress at all depths. In reality this pressure gradient is sufficient to drive the equatorial undercurrent because the fluid at depth feels the pressure force but not the wind stress (Charney, 1960; Gill, 1971). Obviously the layered model misses this effect.

2. The usual finite difference assumption about quadratic terms has been made, that

$$h_1^{-1} \int_{z_1}^{z_2} (\mathbf{u} \cdot \nabla)\,\mathbf{u} \sim \mathbf{u}^1 \cdot \nabla\,\mathbf{u}^1.$$

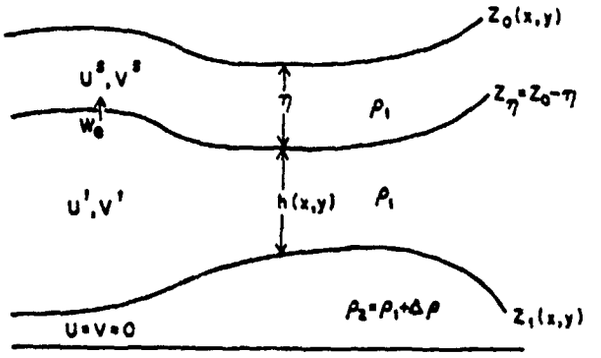


Figure 2. Model with two active layers.

Similar no-motion solutions can easily be found for a multilayer model whether or not the bottom layer is constrained to be motionless: the layer depths may always adjust to reduce the pressure gradient to zero in each subsurface layer. For example, with some dissipation in the system a two layer model with its lower layer not constrained to be motionless will still evolve to such a motionless steady state. (This model does permit a *transient* undercurrent.)

It should be emphasized that such models are not wrong in some simple sense. In fact, the profile of the thermocline depth specified by (2.3) is very close to what is observed at the equator (cf. Gill, 1975, Fig. 3). The difficulty is that the feature of interest is missed by the layered models because they consider only the depth averaged currents within each layer. A correct treatment of the wind stress would introduce it as a boundary condition e.g. $\nu_v u_z = \tau_{wind}$ at the surface, so the velocity cannot vanish at all depths. The vertically *averaged* velocities may vanish. For the example discussed above, this could come about at the equator if the surface flow driven westward by the wind stress were just compensated by the flow at depth driven eastward by the pressure force. In reality, inertial effects act nonuniformly with depth to give a net eastward transport at the equator. This is precisely the mechanism for generating an undercurrent referred to in Sec. 1.

b. Model equations. To capture this essential mechanism we modify the model with a single active layer by dividing this layer into two parts: a surface layer of *constant* depth η and a lower layer of variable depth h (Fig. 2). There is *no density difference* between these two layers and transfer of mass and momentum between the two is permitted. The wind stress is felt directly only by the surface layer; in this sense this layer plays the role of the ocean mixed layer.

Denote the average of a quantity q over the upper active layer by q^s and over the lower layer by q^1; i.e.

$$q^s = \eta^{-1} \int_{z_\eta}^{z_0} q\, dz;\ q^1 = h^{-1} \int_{z_1}^{z_\eta} q\, dz\,. \tag{2.4}$$

The surface $z = z_\eta$ is not a material surface so $dz_\eta/dt \neq w(z_\eta)$; in fact

$$\frac{dz_\eta}{dt} = \frac{d(z_0-\eta)}{dt} = \frac{dz_0}{dt} = w(z_0) \tag{2.5}$$

and

$$w(z_\eta) = \frac{dz_\eta}{dt} + w_e = w(z_0) + w_e \tag{2.6}$$

where w_e is the rate at which fluid must be transferred across the interface $z = z_\eta$ in order to keep the upper layer depth constant. By integrating the continuity equation, $w_z + \nabla \cdot \mathbf{u} = 0$, through the upper layer and using (2.6) we may obtain the relation

$$w_e = w(z_\eta) - w(z_0) =$$

$$\int_{z_\eta}^{z_0} \nabla \cdot \mathbf{u}\, dz = \nabla \cdot \int_{z_\eta}^{z_0} \mathbf{u}\, dz - \{\mathbf{u}(z_0) \cdot \nabla z_0 - u(z_\eta) \cdot \nabla z_\eta\}\,.$$

Since

$$\{\mathbf{u}(z_0) \cdot \nabla z_0 - u(z_\eta) \cdot \nabla z_\eta\} =$$

$$\left\{\frac{dz_0}{dt} - \frac{\partial z_0}{\partial t}\right\} - \left\{\frac{dz_\eta}{dt} - \frac{\partial z_\eta}{\partial t}\right\} = \frac{d}{dt}\{z_\eta - z_0\} - \frac{\partial}{\partial t}\{z_\eta - z_0\} = 0$$

it follows that

$$w_e = \nabla \cdot (\eta \mathbf{u}^a) = \eta \nabla \cdot \mathbf{u}^a\,, \tag{2.7a}$$

so that entrainment balances the divergence in the surface layer. A similar manipulation on the continuity equation integrated from z_0 to z_1 yields

$$\frac{\partial h}{\partial t} = -\nabla \cdot (h\mathbf{u}^1) - w_e\,; \tag{2.7b}$$

the depth of the lower layer is changed either by divergence of fluid within the layer or by exchanges with the layer above.

To derive the momentum equations for the two layers we begin with the momentum equation in the form

$$\mathbf{R} = \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + w\frac{\partial \mathbf{u}}{\partial z} = \frac{\partial}{\partial t}\mathbf{u} + \nabla \cdot (\mathbf{u}\mathbf{u}) + \frac{\partial}{\partial_z}(w\mathbf{u})$$

with $\mathbf{R}$ standing for all the other terms in the equation. Integrating over the upper layer and using Leibnitz's rule

$$\int_{z_\eta}^{z_0} \mathbf{R}\, dz = \left\{\frac{\partial}{\partial t}\int_{z_\eta}^{z_0} \mathbf{u}\, dz - \left[\mathbf{u}(z_0)\frac{\partial z_0}{\partial t} - \mathbf{u}(z_\eta)\frac{\partial z_\eta}{\partial t}\right]\right\}$$

$$+\left\{\nabla \cdot \int_{z_\eta}^{z_0} \mathbf{u}\,\mathbf{u}\, dz - [\mathbf{u}(z_0)(\mathbf{u}(z_0) \cdot \nabla z_0) - \mathbf{u}(z_\eta)(\mathbf{u}(z_\eta) \cdot \nabla z_\eta)]\right\}$$

$$+ w(z_0)\,\mathbf{u}(z_0) - w(z_\eta)\,\mathbf{u}(z_\eta)$$

$$= \frac{\partial}{\partial t}(\eta \mathbf{u}^e) + \nabla \cdot (\eta \mathbf{u}^e \mathbf{u}^e) - \mathbf{u}(z_0)\frac{dz_0}{dt} + \mathbf{u}(z_\eta)\frac{dz_\eta}{dt} + w(z_0)\mathbf{u}(z_0) - w(z_\eta)\,\mathbf{u}(z_\eta);$$

$$\int_{z_\eta}^{z_0} \mathbf{R}\,dz = \frac{\partial}{\partial t}(\eta \mathbf{u}^e) + \nabla \cdot (\eta \mathbf{u}^e \mathbf{u}^e) - w_e\,\mathbf{u}(z_\eta) \tag{2.8a}$$

with the last equality following from (2.6).
In the lower layer

$$\int_{z_1}^{z_\eta} \mathbf{R}\,dz = \frac{\partial}{\partial t}\int_{z_1}^{z_\eta} \mathbf{u}\,dz + \nabla \cdot \int_{z_1}^{z_\eta} \mathbf{u}\,\mathbf{u}\,dz - u(z_\eta)\frac{dz_\eta}{dt} + u(z_1)\frac{dz_1}{dt}$$
$$+ w(z_\eta)\,u(z_\eta) - w(z_1)\,\mathbf{u}(z_1),$$

$$\int_{z_1}^{z_\eta} \mathbf{R}\,dt = \frac{\partial}{\partial t}(h\mathbf{u}^1) + \nabla \cdot (h\mathbf{u}^1\mathbf{u}^1) + w_e\,\mathbf{u}(z_\eta) \tag{2.8b}$$

To close the set of equations (2.7), (2.8) $u(z_\eta)$ must be expressed in terms of the other variables. Applying the requirement that there be no spurious sources or sinks of energy determines that

$$\mathbf{u}(z_\eta) = \tfrac{1}{2}(\mathbf{u}^e + \mathbf{u}^1) \tag{2.8c}$$

Writing **R** explicitly and using (2.7) to go from flux to momentum form gives the final form of the momentum equations.

$$\frac{\partial}{\partial t}\mathbf{u}^e + (\mathbf{u}^e \cdot \nabla)\,\mathbf{u}^e + \frac{w_e}{2\eta}(\mathbf{u}^e - \mathbf{u}^1) + f\,\mathbf{k} \times \mathbf{u}^e$$
$$= -g'\nabla h + \frac{\tau_0}{\eta} - \frac{K}{\eta}(\mathbf{u}^e - \mathbf{u}^1) + \nu_H \nabla^2 \mathbf{u}^e, \tag{2.9a}$$

$$\frac{\partial}{\partial t}\mathbf{u}^1 + (\mathbf{u}^1 \cdot \nabla)\,\mathbf{u}^1 + \frac{w_e}{2h}(\mathbf{u}^e - \mathbf{u}^1) + f\,\mathbf{k} \times \mathbf{u}^1$$
$$= -g'\nabla h - \frac{K_B}{h}\mathbf{u}^1 + \frac{K}{h}(\mathbf{u}^e - \mathbf{u}^1) + \nu_H \nabla^2 \mathbf{u}^1, \tag{2.9b}$$

The stresses at the interfaces, $\tau(z_\eta)$, and $\tau(z_1)$ have been modeled in a simple linear fashion:

$$\tau(z_\eta) = K(\mathbf{u}^e - \mathbf{u}^1);\ \tau(z_1) = K_B\,\mathbf{u}^1. \tag{2.10}$$

In terms of the vertical eddy viscosity ν_v, $\tau = \nu_v\,\partial \mathbf{u}/\partial z$ so we may argue heuristically that

$$K = K_B \approx \nu_v/H_e \tag{2.11}$$

with H_e a characteristic vertical distance between fluid elements in the active layers.

Table 1. Standard Values of Model Parameters.

Parameter	Value	Remarks
τ	.465 dyn cm^{-2}/(gm cm^{-3})	Wind stress per unit mass
g'	.1724 msec^{-2}	"reduced gravity" $g' = g\,(\rho_2-\rho_1)/\rho_2$ for $(\rho_2-\rho_1)/\rho_2 = 1.86 \times 10^{-2}$
ν_H	5.86×10^8 cm^2 sec^{-1}	coefficient of horizontal viscosity
ν_v	15 cm^2 sec^{-1}	coefficient of vertical viscosity
K	1.5×10^{-3} msec^{-1}	interfacial friction parameter; $K = \nu_v/H^*$ for $H^* = 100$ m
K_s	1.5×10^{-3} msec^{-1}	bottom friction parameter; $K_s = K$
X_s	28.6° (3184 km)	zonal extent of the basin
θ_2	15°	basin walls are at 15N and 15 S
θ_s	−15°	
η	25 m	depth of upper layer
H_1	175 m	mean value of the lower layer depth h
β	2.2×10^{-13}m^{-1}sec^{-1}	$\beta = (df/dy)_{y=0} = 2\Omega/R$ where $\Omega = 2\pi$ day^{-1} and R is the radius of the earth

Finally, at all lateral boundaries we impose no-slip boundary conditions

$$\mathbf{u}^e = \mathbf{u}^1 = 0\,. \tag{2.12}$$

c. Choice of parameter values. The values for the model parameters given in Table 1 are by and large typical values for equatorial oceans. Placing the zonal coasts at ±15° makes them sufficiently far from the equator so that their presence has negligible influence on the dynamics in the vicinity of the equator (8S to 8N). The possibility of separating the effects of zonal walls from the equatorial dynamics depends on these dynamics being locally determined; i.e., "trapped" to the equator. That this is the case is borne out by our subsequent analytic investigations (also see Cane and Sarachik, 1976, 1977); it is also evident from the flow field pictures obtained from the numerical calculations (see II). Additional shorter numerical experiments with the zonal walls 20° from the equator differed little near the equator from those with the walls at 15°. The zonal width of the basin is smaller than that of the world's oceans, but is large enough to have a broad interior region where the dynamics may be clearly separated from the dynamics of the meridional boundary layers. The model ocean is taken to be on an equatorial beta plane (e.g. Veronis, 1963a, b) with the Coriolis parameter $f = \beta y$.

Vertical eddy diffusion is to be the principal dissipative mechanism in the model. We choose H_* to be 100 meters (one half of the depth of the active layers) and use (2.11) to relate the K's to the vertical eddy viscosity coefficient ν_v. Knauss (1966) calculated a value of 5 cm^2 sec^{-1} by fitting a parabola to the velocity profile of the undercurrent observed in the Pacific. Williams and Gibson (1974) applied universal similarity and local isotropy assumptions to measurements of small scale temperature fluctuation at 150W and a depth of 100 m. They found values of ν_v of 25 cm

sec^{-1} at the equator and 12 cm sec^{-1} at 1N. Charney (1960) and Charney and Spiegel (1971) found that their models best fit the observed undercurrent for a value of the eddy viscosity in the range 14-17 $\text{cm}^2\ \text{sec}^{-1}$. In the light of this evidence, we use 15 $\text{cm}^2\ \text{sec}^{-1}$ as a standard value for ν_v, feeling some confidence in (at least) the order of magnitude of the choice. The horizontal eddy viscosity is taken small enough to have negligible effect on the interior dynamics. A nonzero value is needed if the boundary conditions (2.12) are to be satisfied.

The presence of the surface layer introduces another parameter, the layer depth η. The numerical value we attach to η will determine how the vertically integrated transport is divided between the two active layers. For example, if $\eta = 25$ m and H, the total depth of the layer, is 200 m, then u^2 is the average zonal velocity in the top 25 m and u^1 is the average zonal velocity in the next 175 m. Their depth-weighted sum $25\,u^2 + 175\,u^1$ is the zonal transport. The choice of the surface layer depth has two effects on the model physics, as may be seen by considering its effect on the transport equations. First, the bottom drag is proportional to the lower layer velocity, whose value will depend on the value of η. This is true even in a linear model (cf. Sec. 3). The second effect is nonlinear, and comes about because we make the modelling assumption that the velocities are independent of depth within each layer. This means that the way we choose to divide up the average velocity affects the size of the nonlinear terms.

Because the choice of the surface layer depth does affect the model physics, we seek a physical basis for determining its value. Unfortunately, the available observational evidence from the world's oceans is not sufficient to help us choose this parameter. We make the choice on theoretical grounds. Consider a shallow homogeneous ocean driven by an imposed wind stress. The ocean is specified to be shallow so that the horizontal component of the Coriolis force may be ignored everywhere. Extra-equatorially, the wind stress is felt in an Ekman layer of depth $D_E = \{2\nu_v/f\}^{\frac{1}{2}}$. Below this boundary layer (and away from the bottom) the dynamics are inviscid and it is transmitted via the boundary layer pumping of the Ekman layer. (See, for example, Charney, 1955; Pedlosky, 1968; or Robinson, 1970 for a detailed account.) As the equator is approached, the Ekman depth D_E increases, becoming infinite at the equator in the absence of additional dynamical balances. We are, however, interested in modelling a parameter range when the wind stress is sufficiently strong and the value of the vertical viscosity sufficiently small so that inertial effects become important in the vicinity of the equator. A measure of these effects in the boundary layer is the Rossby number R_0 based on the boundary layer velocity. For a wind stress per unit mass of magnitude τ the velocity scale in the Ekman layer is given by $U = \tau/(D_E f)$. Then $R_0 = U/(fy) = \tau\,[2\nu_v\beta^2 y^3]^{-\frac{1}{2}}$ so that

$$y = R_0^{-2/3}\,[\tau^2/(2\nu_v\beta^2)]^{1/3} \tag{2.13}$$

The inertial terms will enter into the boundary layer momentum balance (along

with the Coriolis and vertical friction terms) when $R_0 = 0(1)$. As the equator is approached, R_0 increases so that equatorward of some latitude Y_C the inertial effects will prevent the boundary layer from deepening any further. In fact, if the velocities increase toward the equator, we may expect that the boundary layer will get shallower. If we assume that the boundary layer stops deepening when $R_0 = .5$ and use the values in Table 1 (i.e, $\tau = .5$ cm^2 sec^{-2}, $\nu_e = 15$ cm^2 sec^{-1}), we obtain $Y_C = 2°$. The Ekman depth D_E is approximately 25 m at this latitude. [Neither of these is very sensitive to the precise value of R_0 for $R_0 = 0(1)$.] These values agree well with Charney and Spiegel's (1971) calculation for the same parameter values (see their Fig. 1). On the basis of this argument we choose the value $\eta = 25$ m, so that our surface layer will contain the boundary layer to be expected from a continuous model.

3. Formulation of the mathematical problem

To facilitate analytic treatment of Eqs. (2.7), (2.9) we scale the variables as follows:

$$(x,y) = L\,(x',y') \qquad (\mathbf{u}^e,\mathbf{u}^1) = u\,(\mathbf{u}^{e\prime},\mathbf{u}^{1\prime})$$
$$\eta = \hat{H}\,\alpha \qquad w_e = [u\hat{H}/L]\,w_e'$$
$$h = \hat{H}_1 + [U\beta L^2/g']\,h' \qquad \tau = \tau_0\,\tau'$$
$$t = Tt'$$

Here $\hat{H}_1$ is the mean depth of the lower active layer and $\hat{H} = \hat{H}_1 + \eta$.

The velocity scale is related to the wind stress magnitude by $U = \tau_0\,(\hat{H}\beta L)^{-1}$. We take the length and time scales as the baroclinic equatorial ones (e.g., Matsuno, 1966; Blandford, 1966): $L = (c/\beta)^{\frac{1}{2}}$ and $T = (c\beta)^{-\frac{1}{2}} = (\beta L)^{-1}$, where $c = (g'\hat{H})^{\frac{1}{2}}$. These length and time scales are internal scales, picked out by the dynamics of the fluid motions. We assume that the wind stress is a smooth function at these scales and that the dimensions of the basin are large compared with L. (For the values in Table 1, $L = 296$ km, $T = 42.6$ hours and $c = 1.92$ m sec^{-1}.)

Dropping the primes and denoting differentiation with subscripts the scaled equations are:

$$\mathbf{u}_t^e + \epsilon\,\{(\mathbf{u}^e \cdot \nabla)\,\mathbf{u}^e + \frac{w_e}{2\alpha}\,(\mathbf{u}^e - \mathbf{u}^1)\} + y\,\mathbf{k} \times \mathbf{u}^e$$
$$= -\nabla h + \tau/\alpha + A\,\nabla^2\,\mathbf{u}^e - \gamma_I\,(\mathbf{u}^e - \mathbf{u}^1)/\alpha\,, \tag{3.1a}$$

$$\mathbf{u}_t^1 + \epsilon\,\{(\mathbf{u}^1 \cdot \nabla)\,\mathbf{u}^1 + \frac{w_e\,(\mathbf{u}^e - \mathbf{u}^1)}{2\,(1-\alpha+\epsilon\,h)} + y\,\mathbf{k} \times \mathbf{u}^1$$
$$= -\nabla h + A\,\nabla^2\,\mathbf{u}^1 + \frac{\gamma_I\,(\mathbf{u}^e - \mathbf{u}^1)}{1-\alpha+\epsilon\,h} - \frac{\gamma}{1-\alpha+\epsilon\,h}\,\mathbf{u}^1\,, \tag{3.1b}$$

$$w_e = \alpha\,\nabla \cdot \mathbf{u}^e\,, \tag{3.1c}$$

$$h_t + (1-\alpha)\,\nabla\cdot\mathbf{u}^1 + \alpha\,\nabla\cdot\mathbf{u}^2 + \epsilon\,\nabla\cdot(h\mathbf{u}^1) = 0\,. \tag{3.1d}$$

where the following nondimensional numbers have been introduced:

$$\begin{aligned}
&\text{Rossby number} && \epsilon = U/(\beta L^2);\\
&\text{Horizontal Ekman number} && A = \nu_E/(\beta L^3);\\
&\text{Interfacial Ekman number} && \gamma_I = K/(\beta L\hat{H});\\
&\text{Bottom Ekman number} && \gamma = K_B/(\beta L\hat{H});\\
&\text{Nondimensional boundary layer depth} && \alpha = \eta/\hat{H}.
\end{aligned} \tag{3.2}$$

The last three of these numbers, while logically independent as the model is formulated, are all related to vertical friction and so may be related to one another. First, with $K = K_B$ we have $\gamma_I = \gamma$. From the arguments of Sec. 2 we expect η to be on the order of the Ekman depth, η_E, at the edge of the equatorial boundary layer $y = L$ (since $L \sim Y_e$). Now

$$\eta_E = (2\nu_v/f)^{\frac{1}{2}}{}_{y=L} = [2KH^*/(\beta L)]^{\frac{1}{2}} = \gamma^{\frac{1}{2}}\,[2H^*/\hat{H}]^{\frac{1}{2}}$$

where H^* is a characteristic layer depth. As before we take $H^* = \hat{H}/2$, so $\alpha = 0\,(\eta_E/\hat{H}) = 0\,(\gamma^{\frac{1}{2}})$.

We are interested in parameter ranges for which vertical friction is more important than horizontal friction: $A << \gamma, \gamma_I$. We also assume that $\gamma^{\frac{1}{2}} < 0(1)$. (For the values of the parameters given in Table 1. $\alpha = .125$, $A = 10^{-4}$, $\gamma = .011$, $\gamma_I = .1$, and this is the case.[3]) Horizontal friction will be neglected in the interior of the basin, including the equator. As previously mentioned, A must be nonzero to allow the boundary conditions (2.12) to be imposed; if $A = 0$ only the normal component of the transport may be set to zero. Sidewall boundary layers will not be considered further here. Cane, 1975 contains a thorough discussion of boundary conditions and sidewall boundary layers in this model. (See Pedlosky, 1968 or Robinson, 1970 for a discussion of sidewall boundary layers in a continuous, unstratified ocean.)

Since it is the linear dynamics of the model which are to be investigated analytically, we linearize (3.1) by assuming $\epsilon = 0$. It is convenient for this analysis to introduce two new variables:

$$\hat{\mathbf{u}} \equiv (1-\alpha)\,\mathbf{u}^1 + \alpha\,\mathbf{u}^2; \quad \tilde{\mathbf{u}} \equiv \alpha\,(\mathbf{u}^2 - \mathbf{u}^1) \tag{3.3}$$

Then by taking appropriate combiations of (3.1a) and (3.1b) one obtains

$$\hat{\mathbf{u}}_t + y\,\mathbf{k}\times\hat{\mathbf{u}} + E\,\hat{\mathbf{u}} = \tau + A\,\nabla^2\hat{\mathbf{u}} + \frac{\alpha}{1-\alpha}\gamma\,\tilde{\mathbf{u}}\,, \tag{3.4}$$

3. Neglect of A in (3.1a) or (3.4) requires $A<\gamma^{3/2}$; neglect in (3.1b) or (3.5a) requires $A<\gamma^2$, which is only marginally true. The analytic results should be in qualitative agreement with the linear numerical experiments.

$$\hat{\mathbf{u}}_t + y\,\mathbf{k}\times\hat{\mathbf{u}} + \nabla h = \boldsymbol{\tau} + A\,\nabla^2\hat{\mathbf{u}} - \gamma\,(\hat{\mathbf{u}}-\mathbf{u})\,, \tag{3.5a}$$

$$h_t + \nabla\cdot\hat{\mathbf{u}} = 0\,, \tag{3.5b}$$

where

$$E = (1-\alpha)^{-1}\,[\alpha^{-1}\gamma_I + \alpha\gamma] \simeq 0\,(\gamma^4)\,. \tag{3.6}$$

In the absence of bottom friction ($\gamma=0$) then equations become uncoupled: $\hat{\mathbf{u}}$ may be determined from (3.4) alone and $\hat{\mathbf{u}}$,h from (3.5) alone. The quantity $\mathbf{u}$ is the (scaled) vertically integrated mass transport while $\hat{\mathbf{u}}$ is the frictional layer velocity; extra-equatorially ($y>>E$) $\hat{\mathbf{u}}$ is just the Ekman layer transport. From (3.3) we may write

$$\mathbf{u}^1 = \alpha^{-1}\,\hat{\mathbf{u}} + \mathbf{u}^1\,,$$

which says that the surface velocity is given primarily by frictional effects corrected by the interior velocity $\mathbf{u}^1$. Away from the equator $\mathbf{u}^1$ becomes geostrophic. The variables $\hat{\mathbf{u}}$, $\mathbf{u}^1$ are this two-layer model's analogues of the variables used by Philander (1971) in his analysis of the dynamics of a continuous shallow equatorial ocean.

4. The interior circulation: analytic results

a. Frictional velocity. Eq. (3.4) may be used to obtain $\hat{\mathbf{u}}$ since the term involving $\hat{\mathbf{u}}$ is never greater than $O(E^2)$ relative to the retained terms. This equation is first order in time with only a parametric dependence on x and y (with $A=0$), and so may be solved readily for arbitrary wind stress. It is sufficient to treat only a wind stress that is a step function in time switched on at $t=0$. In this case

$$\hat{\mathbf{u}} = [E^2 + y^2]^{-1}\,\{-y\,\mathbf{k}\times\boldsymbol{\tau} + E\,\boldsymbol{\tau}\}\times\{1 - \exp\,[(-E + i\,y)t]\} \tag{4.1}$$

The time scale for the buildup of this component of the current system is clearly $E^{-1} \sim 20$ days for the values in Table 1. For short times [$t << O(E^{-1})$] and points sufficiently near the equator [$|y| << O(t^{-1})$] (4.1) simplifies to $\hat{\mathbf{u}} = t\boldsymbol{\tau}$; i.e. the solution is in the direction of the wind and grows linearly with time. Right at the equator the solution valid for all time is simply $\hat{\mathbf{u}} = \boldsymbol{\tau}E^{-1}\,\{1-e^{-Et}\}$ so that the $\hat{\mathbf{u}}$ at the equator is always in the direction of the wind. Its magnitude is limited by the friction between the two layers and approaches $E^{-1}\,|\boldsymbol{\tau}|$ for times long compared with E^{-1}. Away from the equator $\hat{\mathbf{u}}$ approaches the Ekman wind drift, $-y^{-1}\,\mathbf{k}\times\boldsymbol{\tau}$; it is directed 90° to the right (left) of the wind in the northern (southern) hemisphere.

b. Vertically integrated transport: steady state circulation. The steady state form of the continuity equation (3.5b) allows us to define a mass transport stream function

$$\mathbf{u} = \mathbf{k}\times\nabla\,\psi\,. \tag{4.2}$$

satisfying $\psi = 0$ at the boundaries. A vorticity equation may then be derived from (3.5a):

$$\gamma \nabla^2 \psi + \psi_x - \mathbf{k} \cdot \nabla \times \tau = \gamma \mathbf{k} \cdot \nabla \times \tilde{\mathbf{u}} . \tag{4.3}$$

For the moment we set the right-hand side of (4.3) to zero, reducing it to the Stommel (1948) model for the mass transport stream function. As is well known, this equation admits boundary layers at the zonal boundaries and at the western side of the basin, but not at the eastern side. The appropriate boundary condition for the interior problem is $\psi = 0$ at $x = X_E$, the eastern boundary. Letting $\tau = (\tau^{(x)}, \tau^{(y)})$ the interior solution may be written

$$\begin{aligned} \psi^{(0)} &= -\int_x^{x_E} \mathbf{k} \cdot \nabla \times \tau \, dx , \\ h^{(0)} &= -\int_x^{x_E} [y \, \mathbf{k} \cdot \nabla \times \tau + \tau^{(x)}] \, dx + \int^y \tau^{(y)} (x = X_E) \, dy \end{aligned} \tag{4.4}$$

If the wind stress curl vanishes everywhere (e.g. for constant τ) then (4.4) says that there is no vertically integrated mass transport; the wind stress is balanced by the pressure gradient force.

In the absence of bottom friction the steady state circulation is completely described by (4.1) and (4.4). However, with γ nonzero the right-hand side of the vorticity equation becomes 0(1) in a region $|y| \leqslant O(E)$. Hence there is a region at the equator in which the circulation controlled by the interfacial friction, which itself has no net transport, induces a mean circulation via bottom friction. Letting $\zeta = E^{-2} y$ and writing

$$\psi (x,y) = \psi^{(0)} (x,y) + \pi (x,\zeta) , \tag{4.5}$$

where π is the stream function for this frictionally induced circulation, we note that π depends only on the local winds at the equator. Since near the equator

$$\mathbf{k} \cdot \nabla \times \tilde{\mathbf{u}} \sim -E^{-2} \tilde{u}_\zeta^{(0)} = E^{-2} \{(1 + \zeta^2)^{-1} (\zeta \tau^{(y)} (x,y) + \tau^{(x)} (x,y))\};$$

[cf. (4.1)], the equation for π derivable from (4.3) is

$$\pi_{\zeta\zeta} + a\pi_x = -\tilde{u}_\zeta^{(0)} \tag{4.6}$$

where $a = \gamma^{-1} E^2 = O(1)$. The solution to (4.6) is derived in Cane (1975) and is depicted (for a=1) in Figures 3 and 4 for uniform easterly and southerly wind stresses.

For a zonal wind stress the net transport at the equator is in the direction of the wind. This is, of course, contrary to what is observed for the undercurrent. It says that we must look to other (i.e. nonlinear) effects to explain the undercurrent. For a meridional wind the transport will be in the direction of the wind drift current in both hemispheres. The fluid circuit will be closed by a weak interior transport directed opposite to the wind and a downwind flow in the western boundary layer.

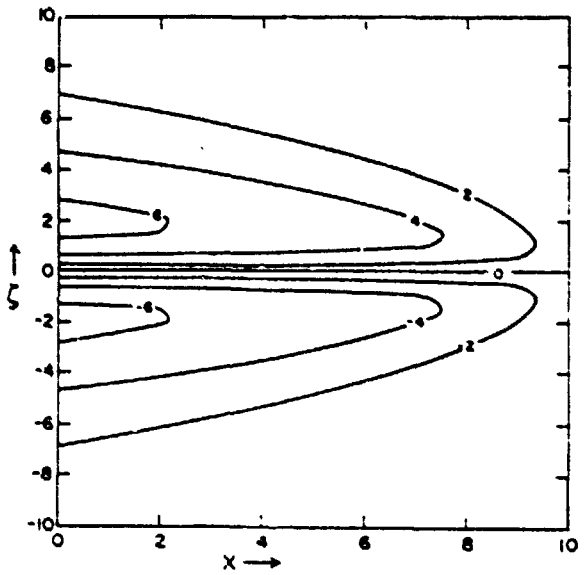


Figure 3. The equatorial boundary layer mass transport stream function $\Pi^{(1)}$ (x,ζ) of Eq. (3.16) for a uniform easterly wind $\tau^{(x)} = -1$, $\tau^{(y)} = 0$.

For any wind stress pattern the flow will be predominantly zonal [$\bar{v}=O(\Sigma\bar{u})$] since flow along the equator is favored. The diffusionlike nature of (4.5) means that the region of frictionally induced transport will broaden from east to west. This description will be compared with the steady state linear numerical results in II.

To summarize, we have found that the steady state interior circulation consists of two parts. The first part, described by (4.1) and (4.3), has a Sverdrup balance everywhere for the transport and essentially a wind drift solution for the boundary layer. The second part, described by (4.6), is important in a region extending about 300 km on either side of the equator. (Note that although $\zeta = 1$ corresponds to only $y =$ 30 km, variables fall off slowly—like ζ^{-1} in some cases.) There is a net transport at the equator in the direction of the zonal wind. Return flow also takes place within this region.

These results may be compared with those of Philander (1971) for a homogeneous ocean continuous in the vertical. For that model, the frictional layer deepens toward the equator and extends throughout the ocean at the equator. The boundary layer in which this happens is embedded in a more diffuse boundary layer in which bottom friction is important. There is a net transport in the direction of the zonal wind in the first of these layers, which is returned in the broader layer. It appears that our modeling assumption, which fixes the boundary layer depth, has the effect of combining these two layers.

c. Vertically integrated transport: time dependent circulation. As with the steady state solution, the time dependent circulation described by (3.5) is best split into

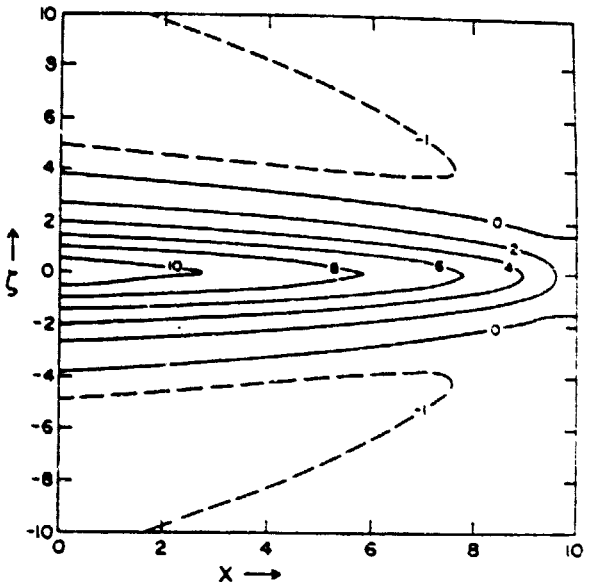


Figure 4. The equatorial boundary layer mass transport stream function $\Pi^{(1)}$ (x,ζ) of Eq. (3.16) for a uniform northerly wind $\tau^{(x)} = 0$, $\tau^{(y)} = 1$.

two parts. By neglecting bottom friction in (3.5a), Eqs. (3.5) become the inviscid shallow water equations on an equatorial beta plane, which may be solved by the methods of Cane and Sarachik (1976, 1977) for all cases of interest here. We will use their results in subsequent descriptions of linear numeral calculations, but for now we only state some of the implications of their work for the spin-up of the present model.

Adjustment toward a final steady state proceeds from east to west and is accomplished by quasi-geostrophic Rossby waves generated at the eastern boundary. The timescale for adjustment (the "setup time") depends on the time it takes for these waves to cross the basin, and so is a linear function of the zonal extent of the ocean. When equatorial Kelvin waves are present (as is usually the case) the adjustment does not proceed smoothly from east to west. Kelvin waves cause mass to oscillate back and forth across the basin so that the final state is not approached monotonically from the east and the adjustment time is lengthened.

Extra-equatorially ($|y| >> E$), the term $\gamma(\tilde{\mathbf{u}}-\tilde{\mathbf{u}})$ in (3.5a) is negligible for all time, but we know from the steady state solution discussed above that it must eventually become important in the equatorial vorticity balance. By rescaling (3.5) it may be shown that frictionally induced transport has a stream function π satisfying the time dependent version of (4.6); viz.

$$\left(\frac{\partial}{\partial t_*} + 1\right)\pi_{\zeta\zeta} + a\,\pi_x = -\tilde{\mathbf{u}}_\zeta^{(0)}\,.$$

As before, $\zeta=E^{-1}y$ while $t_*=\gamma t$ so that this component of the transport evolves on a timescale γ^{-1}: about 200 days for the values in Table 1. The height deviation associated with this transport remains small [$h=O(E)$].

5. Conclusion

The principal purpose of this paper has been the development of a simple model suitable for numerical experiments designed to give insight into the dynamics of the equatorial ocean circulation. There are two layers above the thermocline with *no density difference* between them. The ocean below the thermocline is modeled as a resting layer with a higher constant density. The surface layer is taken to be of constant depth while the depth of the lower active layer is dynamically determined. The two active layers exchange mass (and momentum, energy, etc.) via the vertical velocity at their interface. They are also frictionally coupled. This is the simplest vertical structure that allows a steady state undercurrent.

In our model the wind stress is taken up directly by the surface layer, which thus acts like an Ekman layer in a vertically continuous but homogeneous ocean. In a more realistic model—or the real ocean—it is the surface mixed layer that directly absorbs the wind stress and turbulently mixes the momentum input more or less uniformly throughout the depth of the layer.

Strictly, the wind stress enters the momentum equations as a boundary condition on the vertical shear. By integrating these equations in the vertical a layer model is obtained in which the stress appears as a body force driving the total momentum in the surface. As shown in Sec. 2 the conventional layered model (in which the layer below the surface has a different density and there is no mass exchange between layers) permits a steady state solution in which there is no motion for a curl-free wind stress. Such a model can miss the undercurrent when it should be present. More generally, these models will underestimate inertial effects because the effect of the wind stress is averaged over too great a depth. The mean velocity of the layer is much smaller than the velocity to be expected in the surface mixed (or Ekman) layer. (It is this turbulent surface layer that first becomes nonlinear.) For example, the calculation of O'Brien and Hurlbut (1974), while formally nonlinear, shows nearly linear behavior because the direct effect of the stress is not confined to a sufficiently shallow surface layer.

The linear steady state vertically integrated transport is given primarily by the Sverdrup (1947) balance. For spatially constant winds this Sverdrup transport vanishes. If the stress at the model thermocline is nonzero, there is additional vertically integrated transport in a frictional boundary layer centered on the equator. This layer thickens from east to west. The interior transports are predominantly zonal; a boundary current is required at the western side to close the fluid circuit. A zonal wind stress produces a net transport in the direction of the wind at the equator. This result shows that the *linear* model cannot produce vertically integrated

transport in the same direction as the observed (and model) undercurrent. The nonlinear dynamics must be included to get the correct result.

Two timescales for the evolution of the circulation are revealed by the analysis of the time dependent equations. The time for the boundary layer flow $\tilde{u}$ to become established is $0(\gamma^{-\frac{1}{2}})$ – about 20 days for the parameters used in this model. The transports (or lower layer flow) cannot be establishel until the basin-wide pressure gradients are set up. This requires that mass be moved longitudinally across the basin, a task accomplished by equatorial Kelvin and Rossby waves. The setup time thus dèpends on the zonal extent of the ocean. It also depends on the nature of the forcing: the setup time is longer when Kelvin waves are a significant part of the ocean's response. In all cases the setup time will be at least 100 days for the present model.

Because (at least) the surface velocities in the vicinity of the equator quickly become large, it is clear that the flow becomes nonlinear. The nonlinear dynamics are investigated numerically in II. Those numerical experiments, together with similar experiments on the linear dynamics, will be interpreted in the light of the analytic results reported here.

Acknowledgments. I am grateful to my advisor, Professor J. G. Charney, for his guidance and encouragement. Several valuable conversations with Professor H. M. Stommel are gratefully acknowledged, as are helpful discussions with Professor Moshe Israeli concerning the model formulation. My special thanks to Dr. E. S. Sarachik, for providing invaluable criticisms throughout the course of this work. This work was supported by NASA Grant NGR 22-009-727 to the Massachusetts Institute of Technology.

REFERENCES

Blandford, R. 1966. Mixed Gravity-Rossby waves in the ocean. Deep Sea Res., *13*, 941–960.

Cane, M. A. 1975. A study of the wind-driven ocean circulation in an equatorial basin, Ph.D. Thesis, Masachusetts Institute of Technology, 372 pp.

—— 1978. The response of an equatorial ocean to simple wind stress patterns: II. Numerical results. J. Mar. Res., this issue.

Cane, M. M. and E. S. Sarachik. 1976. Forced baroclinic ocean motions: I. The linear equatorial unbounded case. J. Mar. Res., *34*, 629–665.

—— 1977. Forced baroclinic ocean motions: II. The linear bounded case. J. Mar. Res., *35*, 395–432.

Charney, J. G. 1955. The generation of ocean currents by wind. J. Mar. Res., *14*, 477–498.

—— 1960. Non-linear theory of a wind-driven homogeneous layer near the equator. Deep Sea Res., *6*, 303–310.

Charney, J. G. and S. Spiegel. 1971. Structure of wind driven equatorial currents in homogeneous oceans. J. Phys. Oceanogr., *1*, 149–160.

Düing, K., P. Hisard, E. Katz, J. Krauss, J. Meincke, K. Moroshkin, G. Philander, A. Rybnikov, and K. Voigt. 1975. Meanders and long waves in the Equatorial Atlantic. Nature, *257*, 280–284.

Fofonoff, N. P. and R. B. Montgomery. 1955. The equatorial undercurrent in the light of the vorticity equation. Tellus, *7*, 518–521.

Gill, A. E. 1971. The equatorial current in a homogeneous ocean. Deep Sea Res., *18*, 421–431.
—— 1975. Models of equatorial currents, *in* Proc. Symp. on Numerical Models of Ocean Circulation, Washington, D.C., Nat. Acad. Sci., 364 p.
Knauss, J. A. 1966. Further measurements and observations on the Cromwell Current. J. Mar. Res., *24*, 205–240.
Knox, R. A. 1976. On a long series of measurements of Indian Ocean equatorial currents near Addu Atoll. Deep Sea Res., *23*, 211–221.
Leetmaa, A. 1973. The response of the Somali Current at 2°S to the southwest monsoon of 1971. Deep Sea Res., *20*, 397–400.
Lighthill, M. J. 1969. Dynamic response of the Indian Ocean to onset of the southwest monsoon. Phil. Trans., Roy. Soc., *A 265*, 45–92.
Matsuno, T. 1966. Quasi geostrophic motions in the equatorial area. J. Met. Soc. Japan, *44*, 25–43.
McKee, W. D. 1973. The wind-driven equatorial circulation in a homogeneous ocean. Deep Sea Res., *20*, 889–899.
Munk, W. H. 1950. On the wind-driven ocean circulation. J. Meteorol., *7*, 79–93.
O'Brien, J. J. and H. E. Hurlbut. 1974. An equatorial jet in the Indian Ocean, Theory. Science, *184*, 1075–1077.
Pedlosky, J., 1968. An overlooked aspect of the wind-driven ocean circulation. J. Fluid Mech., *32*, 809–821.
Philander, S. G. H. 1971. The equatorial dynamics of a shallow homogeneous ocean. Geophys. Fluid Dyn., *2*, 219–245.
—— 1973a. The equatorial thermocline. Deep Sea Res., *20*, 69–86.
—— 1973b. Equatorial undercurrent: measurements and theories. Rev. Geophys., *11*, 513–570.
—— 1976. Instabilities of zonal equatorial currents. J. Geophys. Res., *81*, 3725–3735.
—— 1978. Forced oceanic waves. Rev. Geophys. Space Phys., *16*, 15–46.
Robinson, A. R. 1960. The general thermal circulation in the equatorial regions. Deep Sea Res., *6*, 311–317.
—— 1966. An investigation into the wind as the cause of the Equatorial Undercurrent. J. Mar. Res., *24*, 179–204.
—— 1970. Boundary layers in ocean circulation models in Annual Review of Fluid Mechanics *2*, Palo Alto, Annual Reviews Inc., 312 pp.
Stommel, H. 1948. The westward intensification of wind-driven currents. Trans. Amer. Geophys. Union, *29*, 202–206.
—— 1960. Wind-drift near the equator. Deep Sea Res., *6*, 298–302.
Sverdrup, H. D. 1947. Wind-driven currents in a baroclinic ocean, with applications to the equatorial currents of the eastern Pacific. Proc. Nat'l. Acad. Sci., *33*, 318–326.
Taft; B., B. Hickey, C. Wunsch and D. Baker. 1974. The Cromwell Current at 150°W. Deep Sea Res., *21*, 403–430.
Veronis, G. 1963a. On the approximations involved in transforming the equations of motion from a spherical to the β-Plane, part I. Barotropic systems, J. Mar. Res., *21*, 110–124.
—— 1963b. On the approximations involved in transforming the equations of motion from a spherical to the β-Plane, part II. Baroclinic systems, J. Mar. Res., *21*, 199–204.
Veronis, G. and H. Stommel. 1956. The action of variable wind stresses on a stratified ocean, J. Mar. Res., *15*, 43–75.
Williams, R. and C. Gibson. 1974. Direct measurement of turbulence in the Pacific Equatorial Undercurrent. J. Phys. Oceanogr., *4*, 104–108.

Printed in U.S.A. for the Sears Foundation for Marine Research,
Yale University, New Haven, Connecticut, 06520, U.S.A.
Van Dyck Printing Company, North Haven, Connecticut, 06473, U.S.A.

Received: *27 June, 1977*; revised: *22 November, 1978*.

Paper 8

Reprint from Journal of Marine Research, Volume 37, 2, 1979.

The response of an equatorial ocean to simple wind stress patterns: II. Numerical results

by Mark A. Cane[1]

ABSTRACT

The model developed by Cane (1978) is used to study the wind-driven circulation in an equatorial ocean. Simple wind stress patterns are imposed and the model evolution and eventual steady state are calculated numerically. Both linear and fully nonlinear responses are discussed; dynamical arguments are presented to account for the principal features.

The setup time in the model experiments was on the order of several hundred days, with the implication that none of the world's equatorial oceans is in equilibrium with the seasonally varying winds. The Atlantic and Indian Oceans will be close to this equilibrium state while the wider Pacific will not. Nonlinear effects become significant within two weeks. Depending on the form of the wind stress, the inclusion of nonlinearities may either lengthen or shorten the corresponding linear spin up time.

The nonlinear responses to zonal winds included a strong eastward equatorial undercurrent, in agreement with observations. Other aspects of the flow qualitatively resemble the linear response and are largely explicable in terms of wave dynamics. With easterlies, the undercurrent requires a basin-wide zonal pressure gradient: the dynamics are nonlocal and the undercurrent takes several months to reach full strength. With westerlies (as in the Indian Ocean) the subsurface eastward flow is caused by frictional-inertial dynamics that are local and rapid: this undercurrent reaches full strength within one week, consistent with observations at Gan (Knox, 1976).

The nonlinear response to a southerly wind stress is inertially determined and wave concepts are inapplicable. There is an intense eastward "countercurrent" at 3N, entirely independent of the wind stress curl. The zonal mean state is barotropically unstable and westward propagating waves appear. With southeasterlies there is an eastward jet at 4N and undercurrent shifts south of the equator, meandering with longitude. The purely zonal wind cases are stable, suggesting that observed instabilities are due to the equatorial current system as a whole and not the undercurrent itself. Additional theoretical and observational implications of these results are discussed.

1. Introduction

In this paper the response of a bounded equatorial ocean to an imposed wind stress is studied. It is an extension of previous investigations (especially Charney and

1. NASA/Goddard Space Flight Center, Laboratory for Atmospheric Sciences, Greenbelt, Maryland, 20771, U.S.A.

Table 1. Standard values of model parameters.

Parameter	Value	Remarks
τ	.465 dyn $cm^{-2}/(gm\ cm^{-3})$	Wind stress per unit mass
g'	.1724 $msec^{-2}$	"reduced gravity" $g' = g(\rho_2-\rho_1)/\rho_3$ for $(\rho_2-\rho_1)/\rho_3 = 1.76 \times 10^{-4}$
ν_H	5.86×10^6 $cm^2\ sec^{-1}$	coefficient of horizontal viscosity
ν_v	15 $cm^2\ sec^{-1}$	coefficient of vertical viscosity
K	1.5×10^{-3} $msec^{-1}$	interfacial friction parameter; $K = \nu_v/H^*$ for $H^* = 100$ m
K_B	1.5×10^{-3} $msec^{-1}$	bottom friction parameter; $K_B = K$
X_B	28.6° (3184 km)	zonal extent of the basin
θ_T	15°	basin walls are at 15N and 15S
θ_B	−15°	
η	25 m	depth of upper layer
H_1	175 m	mean value of the lower layer depth h
β	$2.2 \times 10^{-11} m^{-1} sec^{-1}$	$\beta = (df/dy)_{\theta=0} = 2\Omega/R$ where $\Omega = 2\pi$ day^{-1} and R is the radius of the earth

Spiegel, 1971) to include zonal variations of the oceanic currents as well as time dependence. The intent is to experiment with a laboratory-like model to gain some insight into equatorial dynamics. We do not attempt to achieve a close mimicry of the real ocean. The linear dynamics were explored rather thoroughly by analytic methods in Cane (1979), henceforth referred to as I. Here we will report on numerical calculations of both the linear and the fully nonlinear responses. Comparisons will be made with the analytic results of I and additional simple analytic models will be invoked to explain some of the phenomena observed in the computations.

The remainder of this section is a brief account of the physical model, while Sec. 2 describes the numerical methods used to obtain solutions. The six sections that follow present the results of experiments with simple wind stress patterns. The final section discusses the implications of these results for the world's oceans.

The physical model. The physical model was formulated in I. That paper also has a lengthy discussion of the choice of values for the governing physical parameters; the results are summarized in Table 1. The model is time dependent and treats fully variations in both the zonal and meridional directions. The vertical structure consists of two layers above the thermocline with the same constant density (Fig. 1). The ocean below the thermocline is taken to be of a higher constant density and to be approximately at rest. The upper of the two active layers is a constant depth surface layer which is acted upon directly by the wind stress. The lower active layer is not directly affected by the wind. Its depth is variable, with the variations being dynamically determined. The two layers communicate via the vertical velocity at their interface, as well as frictionally. Extra-equatorially, this structure is equivalent

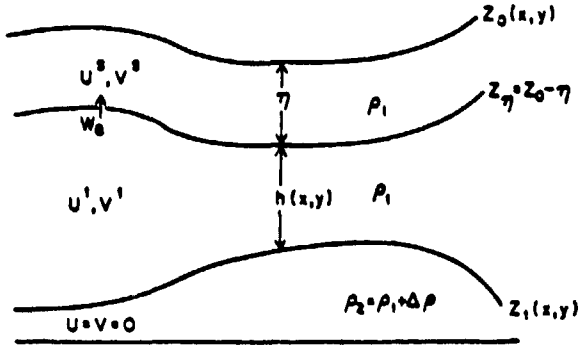


Figure 1. Two layer model.

to a surface Ekman layer and an interior in which the currents are in geostrophic balance.

The layer configuration described above allows for the vertical inhomogeneity that results from the wind stress being felt directly by the ocean at the surface but only indirectly below (e.g., via boundary layer pumping). If the wind stress has no curl, the more usual layered model with each layer having a different density (e.g., Charney, 1955) admits a steady state solution in which each interface tilts in such a way that there is no motion in any of the layers. The present model has the simplest vertical structure that permits a steady state undercurrent.

As in I the equations are nondimensionalized to facilitate our analysis. Letting $\bar{H} = \bar{H}_1 + \eta$, $L = (g'\bar{H}/\beta^2)^{1/4}$, $T = (\beta L)^{-1}$ and $U = \tau_0(\bar{H}\beta L)^{-1}$ we scale the variables as follows:

$$
\begin{aligned}
(x,y) &= L(x'y'), & (\mathbf{u}^2,\mathbf{u}^1) &= U(\mathbf{u}^{2\prime},\mathbf{u}^{1\prime}),\\
\eta &= \bar{H}\alpha, & w_e &= [U\bar{H}/L]w_e',\\
h &= \bar{H}_1 + [U\beta L^2/g']h', & \tau &= \tau_0\tau'\\
t &= Tt',
\end{aligned}
$$

Dropping the primes the governing eqations become (cf. I):

$$
\begin{aligned}
&\mathbf{u}_t^2 + \epsilon\{(\mathbf{u}^2\cdot\nabla)\,\mathbf{u}^2 + \frac{w_e}{2\alpha}(\mathbf{u}^2-\mathbf{u}^1)\} + y\,\mathbf{k}\times\mathbf{u}^2\\
&= -\nabla h + \tau/\alpha + A\,\nabla^2\mathbf{u}^2 - (1-\alpha)\,\gamma_I\,(\mathbf{u}^2-\mathbf{u}^1)\,; \qquad (1.1a)\\
&\mathbf{u}_t^1 + \epsilon\left\{(\mathbf{u}^1\cdot\nabla)\,\mathbf{u}^1 + \frac{w_e\,(\mathbf{u}^2-\mathbf{u}^1)}{2(1-\alpha+\epsilon h)}\right\} + y\,\mathbf{k}\times\mathbf{u}^1\\
&= -\nabla h + A\,\nabla^2\mathbf{u}^1 + \alpha\,\gamma_I\,(\mathbf{u}^2-\mathbf{u}^1)\left[\frac{1-\alpha}{1-\alpha+\epsilon h}\right] - \frac{\gamma\mathbf{u}^1}{1-\alpha+\epsilon h}\,; \qquad (1.1b)
\end{aligned}
$$

$$w_e = \alpha \nabla \cdot \mathbf{u}^e ; \tag{1.1c}$$

$$h_t + (1-\alpha) \nabla \cdot \mathbf{u}^1 + \alpha \nabla \cdot \mathbf{u}^e + \epsilon \nabla \cdot (h\mathbf{u}^1) = 0 . \tag{1.1d}$$

where the following nondimensional numbers have been introduced:

Rossby number	$\epsilon = U/(\beta L^2)$;
Horizontal Ekman number	$A = \nu_H/(\beta L^3)$;
Interfacial Ekman number	$\gamma_I = K/(\beta L \hat{H} \alpha[1-\alpha])$;
Bottom Ekman number	$\gamma = K_B/(\beta L \hat{H})$;
Nondimensional boundary layer depth	$\alpha = \eta/\hat{H}$.

The boundary conditions are that at all lateral boundaries $\mathbf{u}^e = \mathbf{u}^1 = 0$. The parameters are chosen so that the vertical eddy exchanges dominate the horizontal ones. However, as discussed in I, the horizontal eddy viscosity must be nonzero to allow these boundary conditions to be imposed.

Some of the symbols appearing above are defined in Table 1 or Figure 1; the others follow standard oceanographic usage. The model ocean is taken to be on an equatorial β-plane so that the Coriolis parameter is given by $f = \beta y$. Eastward and northward distances are given by the Cartesian coordinates x and y, respectively, with the origin of the coordinate system at the intersection of the equator with the western boundary. The operators ∇ and ∇^2 have their usual two dimensional meanings. The surface layer velocity $\mathbf{u}^e$ has components u^e in the eastward and v^e in the northward direction. Similarly, the velocity of the lower layer, $\mathbf{u}^1 = (u^1, v^1)$. Note that w_e is the surface layer suction: i.e., the rate of exchange between the layers of volume per unit area. This is not the same as the vertical velocity at the interface, $w(z_\eta) \equiv dz_\eta/dt$.

It will prove useful to introduce two quantities defined in I:

$$\tilde{\mathbf{u}} = \alpha \mathbf{u}^e + [1-\alpha+\epsilon h]\, \mathbf{u}^1; \; \hat{\mathbf{u}} = \alpha\, (\mathbf{u}^e - \mathbf{u}^1); \tag{1.3}$$

where $\tilde{\mathbf{u}}$ is the *vertically integrated transport* and $\hat{\mathbf{u}}$ is the frictional boundary layer transport (see I). In mid-latitudes $\hat{\mathbf{u}}$ is the Ekman layer transport. Eqs. (1.1) may be linearized about a resting basic state with a mean lower layer depth H_1 (i.e., $h = \mathbf{u}^e = \mathbf{u}^1 = 0$) by setting $\epsilon \equiv 0$. Then it was shown in I that the governing equations for $\tilde{\mathbf{u}}$ and h are just the linear shallow water equations driven by the wind stress τ. A method for solving the time dependent, inviscid version of these equations has been given by Cane and Sarachik (1976, 1977, 1979) henceforth CSI, CSII and CSIII. The growth of $\tilde{\mathbf{u}}$ results from the wind driving τ and is limited by Coriolis effects extra-equatorially and by the interfacial friction near the equator (see I).

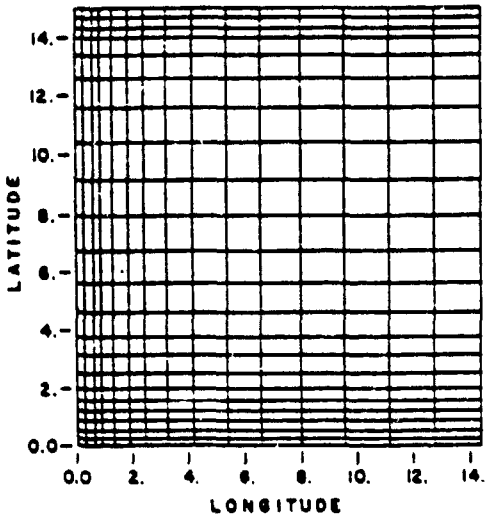


Figure 2. Grid configuration. Only ¼ of the full grid is shown; the grid is symmetric about the equator and the longitude $x = 14.3°$. In this and subsequent figures latitude and longitude are given in degrees (1 degree ≈ 111 km.)

2. Numerical methods

The methods used for the numerical solutions of (1.1) are described in detail in Cane (1975). The salient features will be discussed briefly here.

a. Grid configuration. An unstaggered variable size grid mesh is employed to allow increased resolution for the sidewall boundary layers and along the equator. The grid points are shown in Figure 2. Following a suggestion of M. Israeli (Orszag and Israeli, 1974) the location x_i of the ith point in physical space is determined by solving an equation of the form

$$x_i^* = a + bx_i + \sum_j \alpha_j \tan^{-1}\left[\beta_j^{-1}(x_i - \gamma_j)\right] \tag{2.1}$$

so that the points x_i^* in the "computational space" are equally spaced; i.e. (2.1) is solved for the x_i that gives $x_i^* = (i-1)\Delta x$. Many of the figures (e.g. Fig. 6) are plotted in the computation coordinate x^* rather than the physical coordinate x in order to increase the visual area devoted to regions of interest (e.g. the equator). In (2.1) β_j is (a measure of) the thickness of the boundary layer at $x = \gamma_j$ and α_j is a weighting factor determining how many mesh points will be in this boundary layer.

b. Time differencing. Time marching is accomplished with the N-cycle scheme of Lorenz (1971) with $N = 4$. This scheme is second order in time (fourth order for linear equations). In the calculations reported here we use a timestep of $(4\Omega)^{-1} \sim$ 1 hour.

c. Spatial differencing. All derivatives are approximated by centered fourth order finite differences in the interior and by third order differences at the points on and immediately adjacent to the boundary. Since the spacing at the boundary is smaller than in the interior, this does not degrade the overall accuracy. The finite difference approximation D is made in the computational space where the grid points x_i^* are equally spaced. D may be related to the desired derivative $\partial f/\partial x$ in the physical space by

$$\frac{\partial f}{\partial x} = \frac{dx^*}{dx}\frac{\partial f}{\partial x^*} \sim \frac{dx^*}{dx} Df$$

where dx^*/dx is calculated (analytically) from (2.1).

The equations are differenced in a conservation form; that is, the spatial differencing is such that in the absence of sources and sinks momentum, energy and mass would be conserved. The exception to this is that the treatment of boundary points does not identically conserve these properties. Tests showed the leakage of mass and energy at the boundaries to be negligibly small (the order of computational round-off errors).

d. Gravity wave treatment; smoothing. The use of conservation forms does not prevent short wave length computational modes (e.g. "checkerboard instabilities") from growing to noticeable size. Physically, one expect gravity waves to adjust the flow to a more slowly varying (e.g. geostrophic) balance by propagating away such small scale disturbances. Mesinger (1972) has pointed out that the usual numerical treatments of gravity waves fails to couple the grid points properly. This accounts for the disparity between their physical and numerical roles.

The simplest set of linear equations describing gravity waves is

$$u_t = -g\frac{\partial}{\partial x}h; \quad h_t = -H\frac{\partial}{\partial x}u\,. \tag{2.2}$$

The finite difference analog is obtained by replacing $\partial/\partial x$ by a finite difference approximation D_x:

$$u_t = -gD_x h; \quad h_t = -HD_x u\,. \tag{2.3}$$

From (2.2) a single equation in h is readily derived:

$$h_{tt} = gHh_{xx} \tag{2.4}$$

The analog obtainable from (2.3) is

$$h_{tt} = gHD_x(D_x h)\,. \tag{2.5}$$

Note that the second derivative in (2.4) has been replaced by two successive first derivative approximations D_x^2 instead of a second derivative approximation D_{xx}. For second order centered differences and a two grid point wave, $h_i = h(x_i) = (-1)^i$, the right-hand side of (2.5) is zero, so that (2.5) fails to propagate the wave. On the other hand $D_{xx}h_i = -4\Delta x^{-2}h_i = 0$.

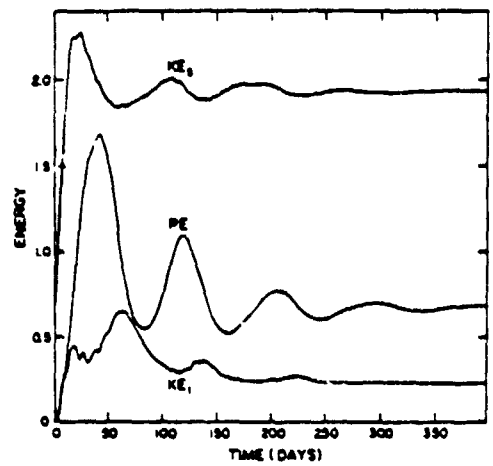


Figure 3. Evolution of energies in the linear response to an easterly wind. Potential energy (PE), and kinetic energy in the surface layer (KE_s) and lower layer (KE_1) for the equatorial region (5.6S to 5.6N) are shown. The unit of energy is $10^3\ Jm^{-2}R^2$, where $R = 6367$ km is the radius of the earth.

Our method of treating gravity waves amounts to replacing the D_x^2 operator in (2.5) with D_{xx} in order to retain the proper coupling between successive mesh points. (The implementation is straightforward but the details are lengthy and will be omitted here; see Cane, 1975, Appendix B.) Equivalently, it amounts to adding a smoother $S_x = gH\ (D_{xx}-D_x^2)$. Since D_{xx} and D_x are the 4th order so is S_x and the formal order of accuracy of the equations is unchanged. Gravity waves in two dimensions require adding operators S_x to the u, S_y to the v and S_x and S_y to the h equations. In addition, it was necessary to add smoothers in the cross stream direction (i.e. S_y to the u and S_x to the v equations) because the horizontal viscosity used in the computations is too small to suppress two-grid point waves where the mesh spacing is large.

3. Linear response to a uniform easterly wind

The model equations are linearized by setting $\epsilon = 0$ in Eqs. (1.1); all model parameter values are as given in Table 1 with a uniform easterly wind stress switched on at $t = 0$. The linear dynamics may be understood rather thoroughly in terms of the analytic results in I and CSI, II, III. [CSIII, Sec. 5 has a detailed account of the early stages of spinup for the shallow water equations.] We present the linear numerical results here for the considerable light they shed on the nonlinear dynamics to be discussed in the following sections; our discussion will be focused accordingly.

The surface flow $\mathbf{u}^s$ is approximated well by the boundary layer velocity $\tilde{\mathbf{u}}$ [see (1.3)]. The latter evolves to a steady state on the frictional timescale of 20 days (see Fig. 3). Extra-equatorially, it approaches the Ekman wind drift solution; flow

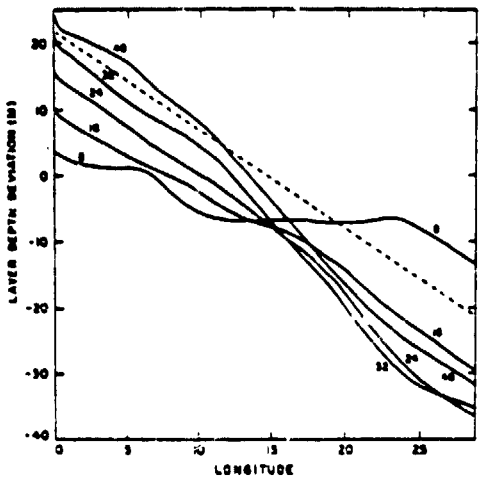


Figure 4. Sections of layer depth deviation, h, along the equator at 8 day intervals in the linear response to an east wind. Each solid curve is labelled with the time in days. The dashed line is the steady inviscid solution (3.2).

is poleward in both hemispheres, resulting in intense upwelling at the equator. Right on the equator the flow is westward, in the direction of the wind, with the turning from westward to poleward occurring within a degree of the equator in a boundary layer in which interfacial friction (γ_i) is important. Figure 3 shows that the lower layer currents are much weaker than those in the surface layer. The linearized dynamics do not allow vertical advection of momentum and the frictional and pressure forces are not capable of generating strong currents.

The time dependent solution for the vertically integrated transports and the layer depth h may be found by the methods of CSI,II. It has $\hat{u}$ and h symmetric and $\hat{v}$ antisymmetric about the equator. In the absence of boundaries, this part of the model response would consist of inertia-gravity waves together with functions of the form

$$(\hat{u},\hat{v},h) = (U(y)t,\, V(y),\, H(y)t) \tag{3.1}$$

(These functions are depicted in CSI Figure 8a for the wind stress $\tau^{(x)} = 1$, which is just the negative of the present case.) U and H are equatorially confined, while V goes to zero at the equator and approaches $-\tau^{(x)}/y$ as y increases. Most of the energy put in by the wind goes into (3.1); relatively little goes into the inertia-gravity waves.

We anticipate that (with the exception of the frictional equatorial boundary layer described in I and discussed further below) the final state will be one in which there is no motion and the surface tilts up uniformly from east to west to balance the wind stress:

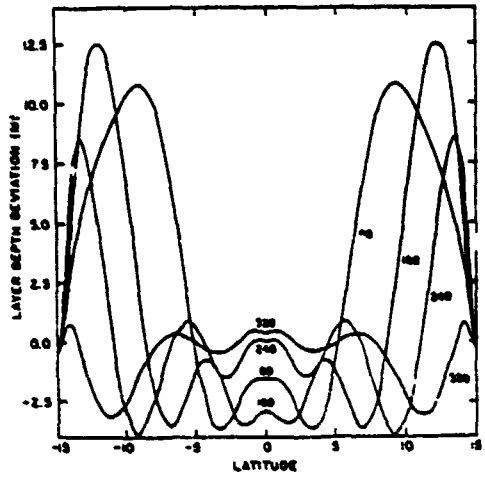


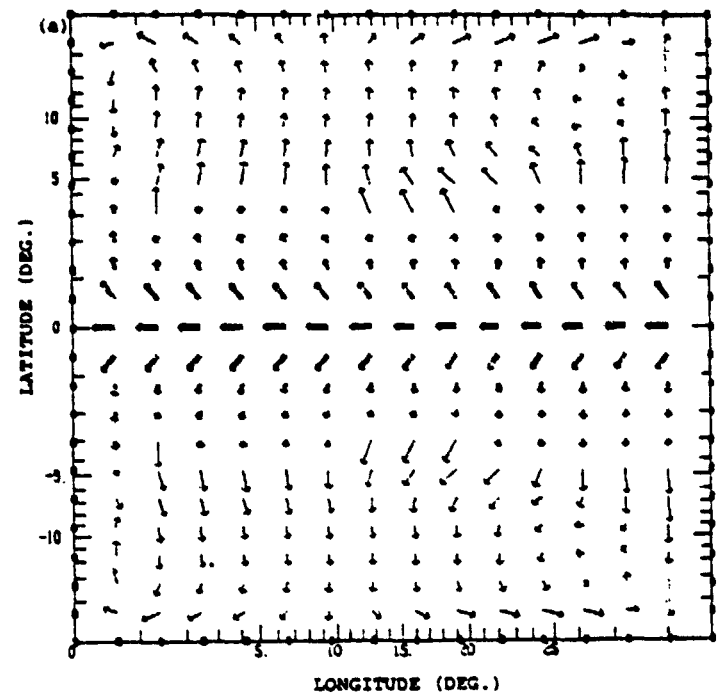
Figure 5. Sections of layer depth deviation, h, along the central longitude $x = 14.3°$ at 80 day intervals in the linear response to an easterly wind. Labels give the time in days.

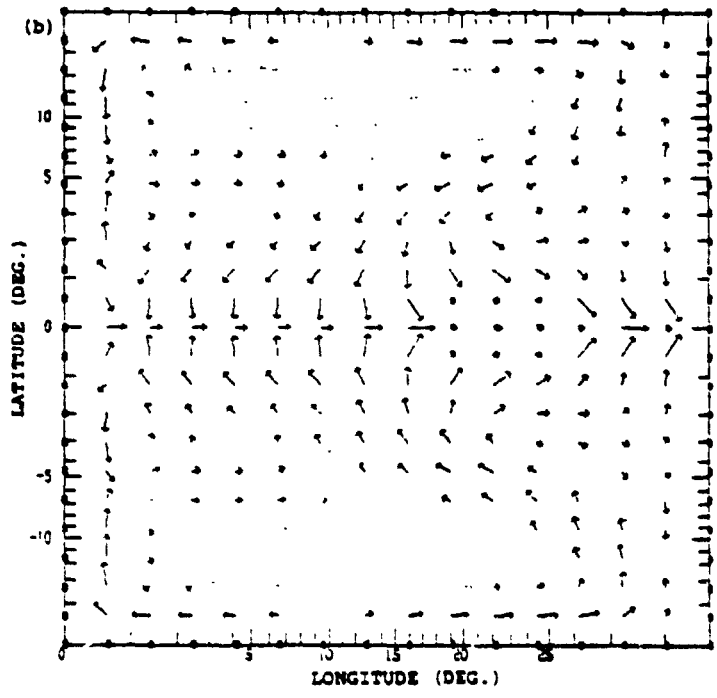
$$\bar{u} = \bar{v} = 0; \quad h = (x - X_E/2)\,\tau^{(x)}\,. \tag{3.2}$$

The adjustment to a final state will be accomplished by the boundary effects on the secularly growing part of (3.1). [See CSII and CSIII for a detailed discussion.] Both the eastern and western boundaries participate actively in the adjustment process because the presence of the eastward travelling Kelvin mode makes it impossible for the adjustment to proceed solely from the east. (Sec. 6 offers a contrasting case.) The Kelvin wave has negligible amplitude extra-equatorially, influencing that region only via the modes generated when it is reflected at the eastern end of the equator. Hence, all the extra-equatorial adjustment proceeds from east to west. The more equatorially confined modes travel the most rapidly so the equatorial region adjusts more rapidly.

The final state is not approached monotonically; for example, the gradient of h tends to "overshoot" its final value. This is evident in the energy curves (Fig. 3) with the most striking oscillations occurring in the potential energy. These oscillations are related to the reflections from the basin walls described in CSII. That analysis shows that the adjustment should have a periodicity of $4T$, where T is the time for a Kelvin wave to cross the basin. The additional $3T$ is the crossing time for the $n = 1$ Rossby wave generated by the Kelvin wave at the eastern boundary. Dimensionally T is 19.8 days so the predicted $4T$ periodicity is in excellent agreement with the 80 period observed in Figure 3.

Figure 4 shows profiles of the layer depth at the equator at various times. If the height were set up to balance the wind stress (3.2), the layer depth profile would be a straight line 21.8 m below (above) the mean value at the eastern (western) side (the dashed line in Fig. 4). During the course of the adjustment the profile tends to


(a)
LATITUDE (DEG.)
LONGITUDE (DEG.)


(b)
LATITUDE (DEG.)
LONGITUDE (DEG.)

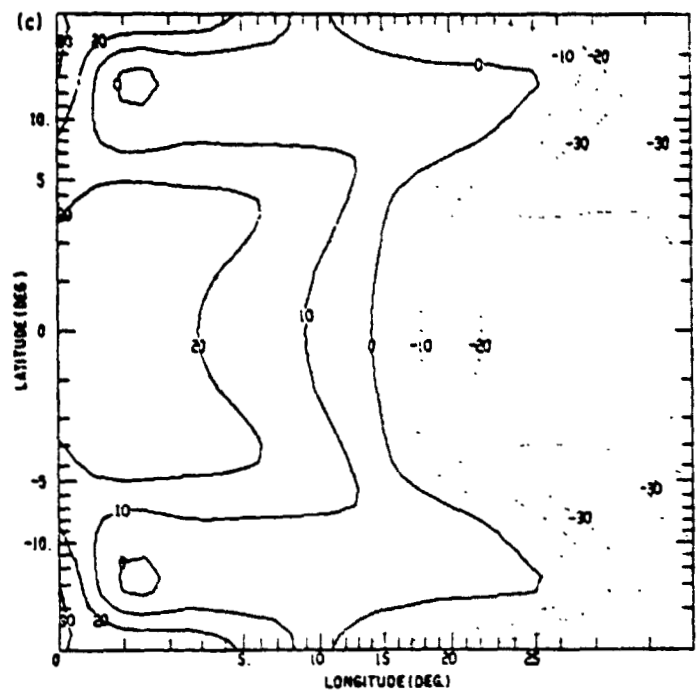


Figure 6. Linear response to an easterly wind at day 40. Note that the coordinates are stretched. (a) Surface layer velocity, u'. Dotted arrows: 0-2 cm s^{-1}; full arrows: 2-20 cm s^{-1}; double arrows: 20-200 cm s^{-1}. (b) Lower layer velocity, u^1. (c) Layer depth deviation, h. Contour interval is 10 m.

be below this final value everywhere. The profile at 8 days shows a flat center section in which the boundary influences have not been felt; here h is decreasing in accordance with the unbounded solution (3.1). At this time, the Kelvin mode generated at the western boundary has propagated 12° into the basin. This is evident in the sloping region at the western side of the basin. Near $x = 0$ there is also evidence of the effect of the boundary trapped modes. At the eastern side there is another sloping piece to the profile extending 4° into the basin. This is due to the Rossby modes generated at the boundary; the fastest of these (the one with meridional index $n = 1$) would have propagated 4° at this time.

At 15 days the two boundary influences meet at a point three-quarters of the way across the basin. Until this time, the magnitude of the westward zonal transport at this point has increased according to (3.1). Hereafter the slope of the height field at all longitudes on the equator will be up toward the west, thus reducing the zonal acceleration. In fact, it is evident from Figure 4 that by 24 days this gradient is generally sufficient to balance the wind stress so that the magnitude of the westward transport will no longer increase. The slopes at the eastern side become steeper than what is required to balance the wind stress so that the transport here becomes

eastward and the layer deepens. This region of eastward flow is behind the front formed by the Rossby modes which are the reflections of the first Kelvin mode to cross the basin. The region propagates out from the eastern boundary beginning at day 20.

Consideration of the zonal transports at the equator indicates that the *equatorial region* takes approximately 250 days to approach its final steady state; this time scale agrees with the energy diagram, Figure 3. At each longitude the steady state layer depth profile is a straight line independent of latitude [cf. (3.2)]. At all longitudes this is attained more rapidly close to the equator. Near the eastern wall h is close to its steady value at all latitudes after 300 days. At the center of the basin the profile still shows some deviation from the expected final value $h = 0$ poleward of about 7° (Fig. 5). There is a strong tilt at the northern and southern walls to geostrophically balance the boundary jets present there [cf. CSIII]. Profiles near the western boundary are even further from their steady state values. As was previously mentioned, analytic theory predicts that the pattern of adjustment will be from east to west and from the equator poleward.

At 40 days (Fig. 6) the layer depth changes are largely confined to the boundaries, an area within 7° of the equator, and the eastern side. Some of the slowly propagating Rossby modes generated at $t = 0$ already have moved away from the eastern boundary. This is clear from the extra-equatorial flow pattern. Behind (i.e., to the east of) this region of equatorward flow in the lower layer (Fig. 6b) there is a region of poleward flow due to the reflection of the Kelvin modes which arrived at the eastern side at day 20. The lower layer flow is eastward at all longitudes on the equator. This transient "undercurrent" is weak ($u^1 < .10$ m sec^{-1}) and narrow, with a frictionally determined half-width of less than 50 km. The direct cause of this eastward flow at depth is the pressure gradient force (Charney, 1960).

By 397 days (Fig. 7) the model ocean is close to a steady state everywhere in the basin with the exception of the northwest and southwest corners. The upper layer is given primarily by a wind-drift-frictional balance [see I, Eq. (4.1)]. There is strong upwelling along the equator with fluid being returned to the lower layer in downwelling regions at the northern and southern boundaries, the western end of the equator, as well as the latitudes near the equator ($\approx \pm 1°$) where v^1 decreases rapidly. There are no exceptionally fast boundary currents. The "undercurrent" maximum velocity is only .2 m sec^{-1} and its half-width is only .5°, numbers determined by the vertical eddy viscosity.

In general, the lower layer flow has an equal and opposite mass flux so that the vertically integrated transport is zero. However, near the equator there is a region of net transport evident in Fig. 7b. This was predicted in I: this net transport is induced by the friction at the bottom of the lower active layer; without this bottom friction the transport would be zero everywhere. As predicted, this circulation occurs in an equatorial boundary layer which thickens from east to west, with pre-

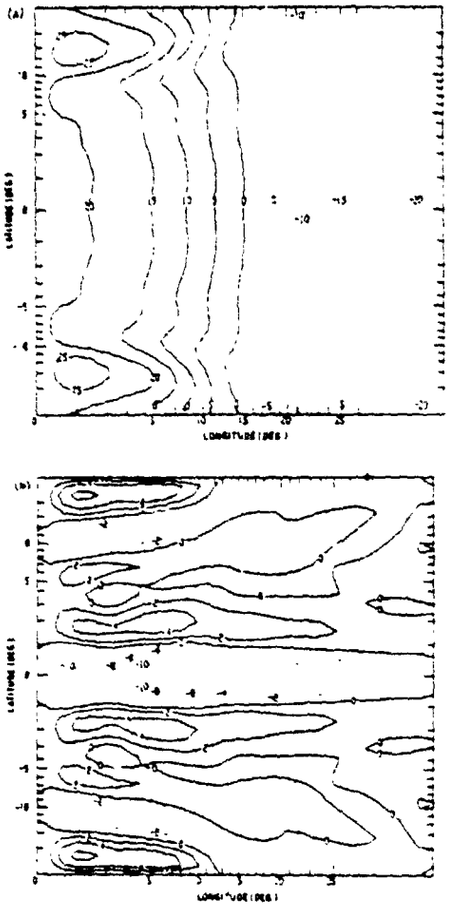


Figure 7. Linear response to an easterly wind at day 397. (a) h. Contour interval is 5 m. (b) Zonal transport u. Contour interval is 2 m^2 sec^{-1}.

dominately zonal transports increasing toward the west. Transport at the equator is in the direction of the wind, while an important signature of the observed undercurrent is that the vertically integrated transport at the equator be large and eastward. Nonlinear dynamics must be included to obtain a realistic undercurrent.

4. Nonlinear response to a uniform easterly wind

In this section we consider the nonlinear response to a uniform east wind. Except for the nonlinearity the governing parameters are identical to those in the previous section. Among the simple wind stress patterns this is the one most closely related to the undercurrent in the Atlantic and Pacific Oceans. We first give a narrative description of the numerical experiment and then consider the governing dynamics.

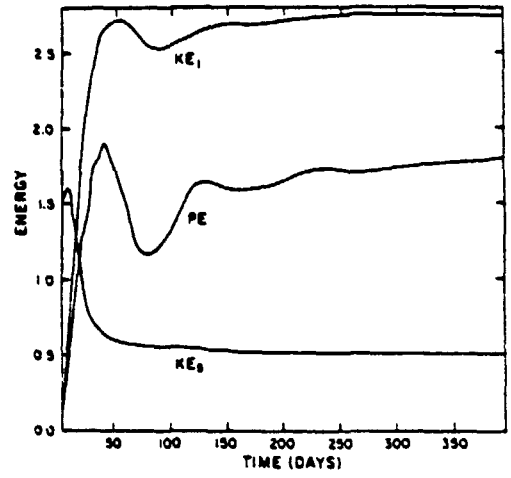


Figure 8. Evolution of energies in the equatorial region for the nonlinear response to an easterly wind. Like Fig. 3. Initial time is 5 days.

a. Evolution of the flow. The energies (Fig. 8) exhibit striking differences from those for the linear response to the same forcing (Fig. 3). The inclusion of vertical momentum advections results in much greater kinetic energy in the lower layer while the surface layer kinetic energy is smaller in the nonlinear case. The upper layer kinetic energy reaches a peak after 9 days, after which it falls off rapidly until day 40.

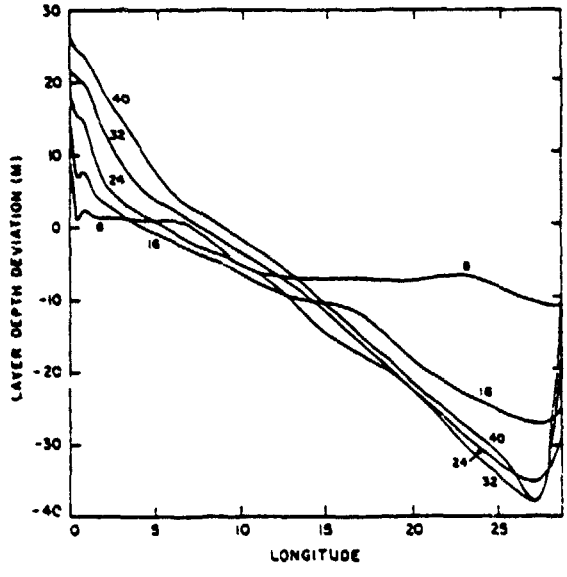


Figure 9. Sections of h along the equator at 8 day intervals in the nonlinear response to an easterly wind. Like Fig. 4.

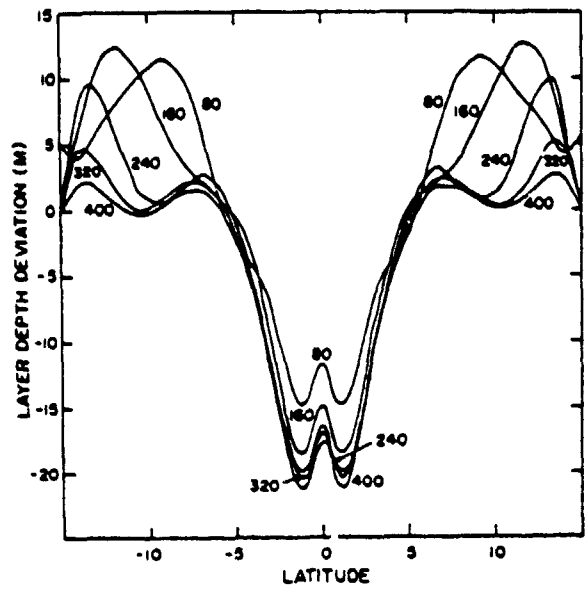
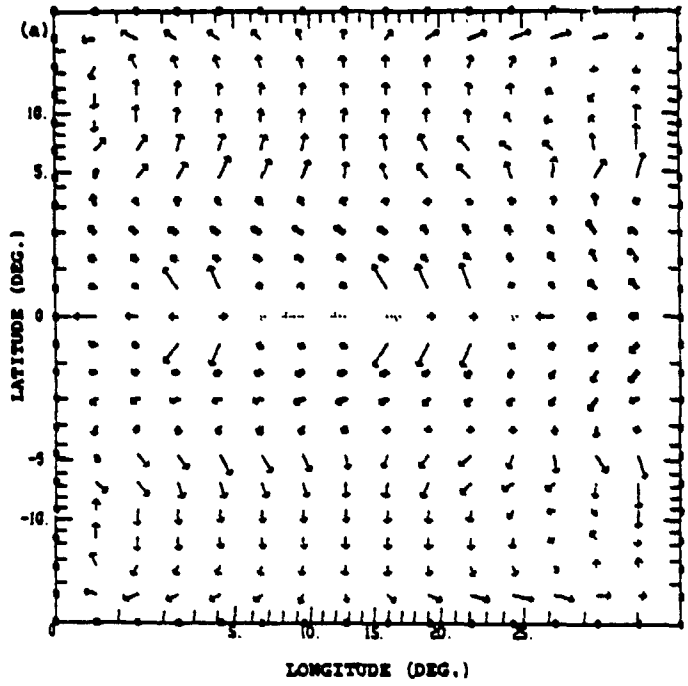
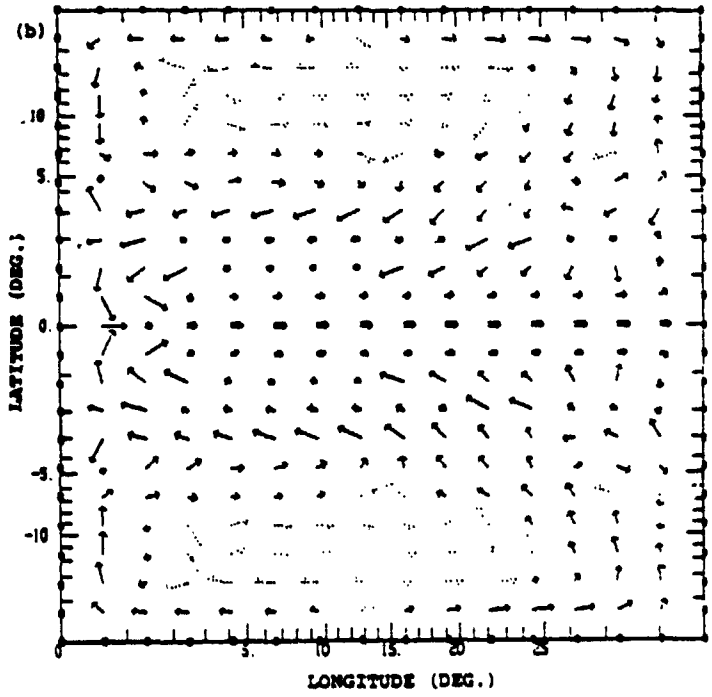


Figure 10. Sections of h along the central longitude $x = 14.3°$ at 80 day intervals in the nonlinear response to an easterly wind. Like Fig. 5.

During this initial 40 day period the potential energy and lower layer kinetic energy both rise to a peak. The loss of surface layer energy to the lower active layer via vertical advections is not fully compensated by the transfer of energy from the winds to the ocean. There is a strong contrast between the linear and nonlinear responses in the oscillations in the energy curves. (These oscillations may be used as an index of the tendency to "overshoot" the final adjusted state as the flow evolves.) The addition of nonlinear effects has damped the tendency to oscillate about a final steady state. Furthermore, the nonlinear steady state is one with potential energy close to the maximum value the system attains in the course of its evolution, while the linear steady state is close to a minimum.

We can see how these differences arise by considering sections of layer depth h at the equator. Figures 9 and 4 show h at the equator for the first 40 days in the nonlinear and linear cases respectively. Throughout this period there is little difference in the two cases, except at the eastern side. Recall that over the first 20 days, the linear response built up a steeper height gradient than was needed to balance the wind stress. After that the reflection of the first Kelvin mode to arrive from the western side caused the layer depth to decrease. The nonlinear case is similar for the first 20 days, but the eastern boundary response to the first signals arriving from the western side is quite different. The slope of h remains steep throughout the basin with a strong boundary layer forming at the eastern wall. Within this narrow (1.5° wide) layer the transports impinging on the eastern wall are turned to the north and south.


(a)
LATITUDE (DEG.)
LONGITUDE (DEG.)


(b)
LATITUDE (DEG.)
LONGITUDE (DEG.)

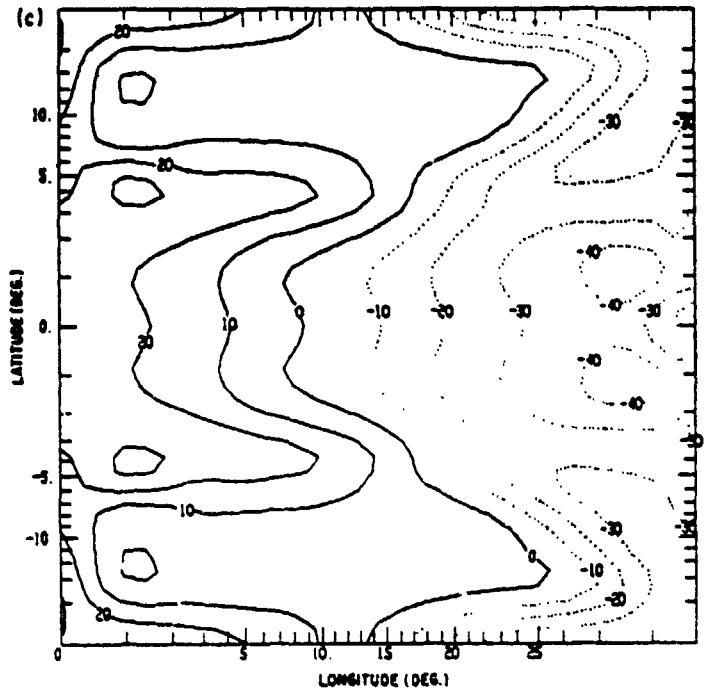
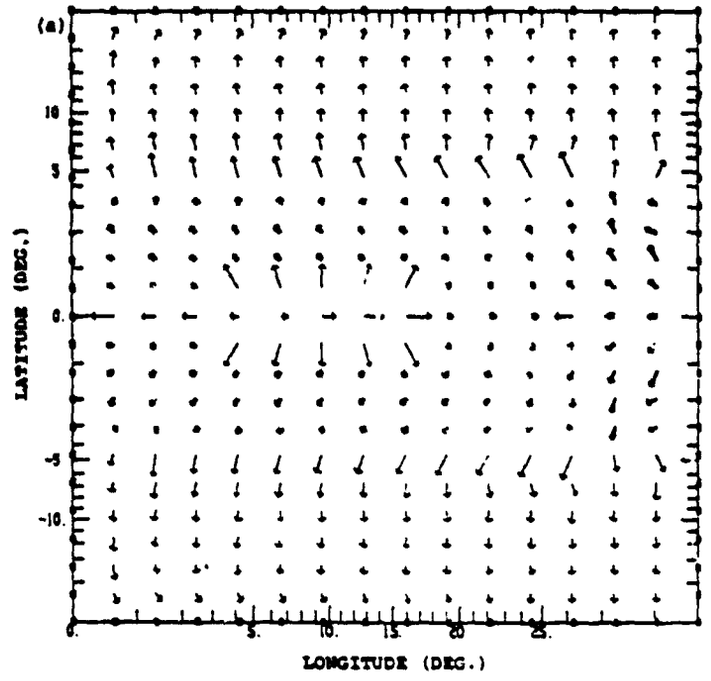


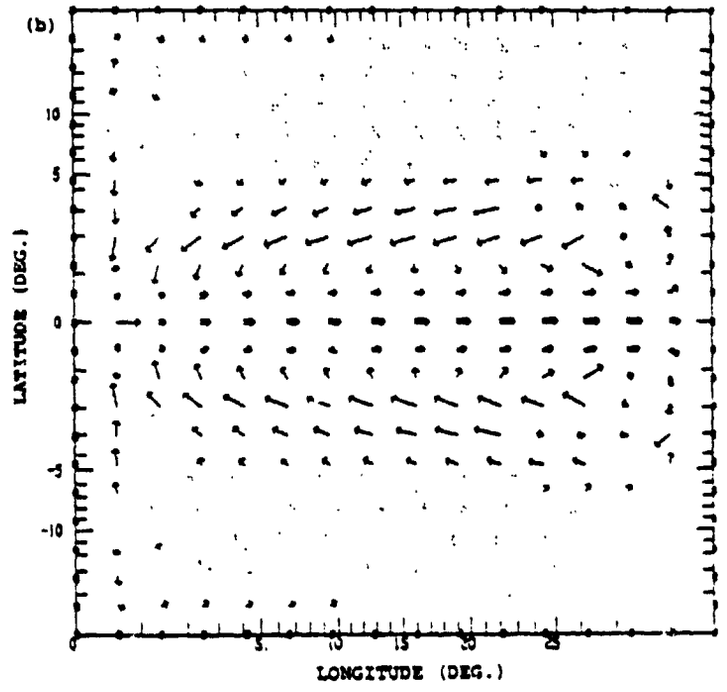
Figure 11. Nonlinear response to an easterly wind at day 40. Like Fig. 6. (a) u' (b) u^1 (c) h. Contour interval is 10 m.

We might say that the nonlinear dynamics respond to the incoming currents at the equator by forming an inertial boundary layer whereas the linear response is a reflection. The Rossby modes which comprise the reflection of the Kelvin mode in the linear theory all propagate too slowly to the west to escape from the boundary in the face of the fast eastward current that exists at the equator. Therefore, they are trapped at the eastern wall on the equator and a boundary layer forms. The layer depth profile evident at day 40 persists thereafter, with its minimum becoming more pronounced with time. After 160 days it varies little.

By 16 days the nonlinear transports along the equator are everywhere eastward, a feature which persists thereafter. The (frictional) linear transports eventuall · became westward everywhere, but they took on the order of 200 days to do so at all longitudes. After one month has elapsed the transports in the nonlinear case are almost an order of magnitude larger. As discussed below, the differences may be explained in terms of the inclusion of relative vorticity in the nonlinear vorticity balance.

Thus far we have considered the solution at the equator only. Figure 10 shows a meridional section of h at the center of the basin. The greatest difference from the corresponding linear section (Fig. 5) is the deep trough within 5.6° of the equator.


(a)
LATITUDE (DEG.)
LONGITUDE (DEG.)


(b)
LATITUDE (DEG.)
LONGITUDE (DEG.)

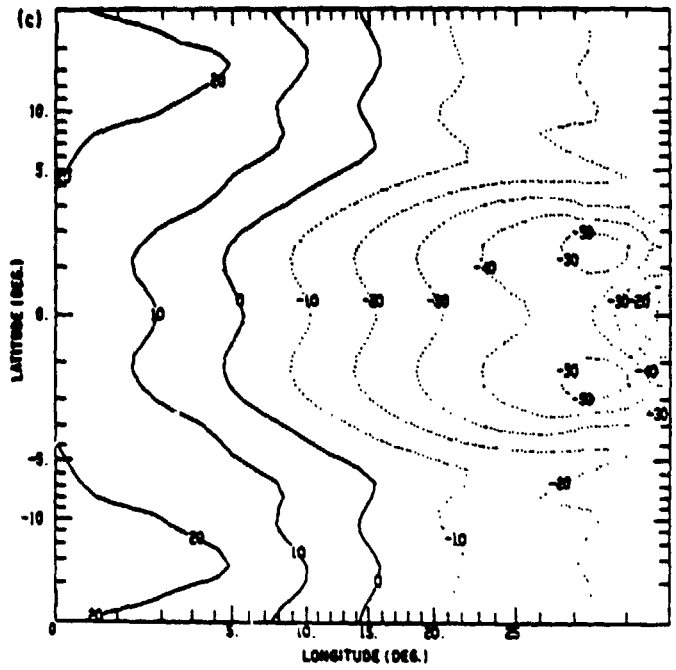


Figure 12. Nonlinear response to an easterly wind at day 398. Like Fig. 7. (a) u' (b) u^1 (c) h. Contour interval is 10 m.

This trough is symmetric about the equator; h slopes downward from 5.6°N to 1.2°N to geostrophically balance the westward current in the lower layer at those latitudes. From 1.2° to 0° it slopes upward to geostrophically balance the undercurrent. By 160 days this trough is close to its final shape. This time scale for adjustment agrees with that given by the energies and the vertically integrated transport at the equator. Within this equatorial region there is some tendency for the adjustment to occur soonest at the eastern side, particularly poleward of about 3°. Poleward of 5.6° the adjustment clearly proceeds from east to west and is more rapid the nearer to the equator one is. The time scales for this extra-equatorial process are comparable to those for the linear case, although h shows some influence from nonlinear effects at all latitudes.

Figure 11 shows the flow at 40 days. The surface layer currents near the equator are more zonally oriented than in the linear case (Fig. 6). This is a result of the fact that parcels tend to turn anticyclonically as they move poleward from the equator and lose relative vorticity to compensate the gain in planetary vorticity. The effects of upwelling, vertical friction and the pressure gradient have reduced the surface layer zonal velocity to near zero at the equator, except at the sidewall boundaries. Vertical advection of eastward momentum from the lower active layer is the most important factor in bringing about this weak surface flow. The maximum undercur-

rent velocity is now above .8 m sec^{-1} and occurs near the eastern boundary. At 14 days this maximum was only .3 m sec^{-1}; by 80 days it is over 1 m sec^{-1}, very close to its final steady value.

At 398 days (Fig. 12) the steady state solution is closely approximated everywhere in the basin with the exception of the northwest and southwest corners. The equatorial region resembles day 40. Poleward of 5° the flow is essentially linear: the surface layer flow is the wind drift solution given by the linear theory while the subsurface flow combines with this to give approximately zero vertically integrated transport. Many of the prominent features near the equator in the interior are in good agreement with the y-z plane calculation of Charney and Spiegel (1971). (See their Figs. 1b and 2b.) Specifically, in both our calculation and theirs the halfwidth of the undercurrent is about 1° and the mean undercurrent velocity is about .80 m sec^{-1}. Both have eastward flow at the surface; as noted above this is primarily due to the strong upwelling at the equator. Eastward flow at the surface with easterly winds has been observed in the Pacific by Taft *et al.* (1974). Both calculations show the strongest westward subsurface flow (on the order of 10 cm sec^{-1}) between 2° and 3°.

b. Basic dynamics. The principal features of the flow may be explained qualitatively by considering the vorticity balance, as in Fofonoff and Montgomery (1955). (Also see Charney, 1960, and Charney and Spiegel, 1971). The easterly wind produces a poleward Ekman drift in the surface layers. This requires upwelling at the equator and therefore an equatorward flow at depth. Parcels moving toward the equator lose planetary vorticity. If we assume that total vorticity is approximately conserved these parcels must acquire relative vorticity as they approach the equator resulting in an eastward flow there. A simple calculation shows that a parcel originating at a latitude y_0 with approximately zero relative vorticity and zero zonal velocity has an eastward velocity of approximately $\beta y_0^2/2$ at the equator. For undercurrent velocities of .75 to 1.00 m sec^{-1} y_0 is between 2.5° and 3°; this is consistent with our calculation. A similar line of reasoning may be used to determine the position of the westward currents. Fluid parcels in the undercurrent that reach the eastern side are turned poleward in narrow boundary currents. As they travel away from the equator they gain planetary vorticity. In order to approximately conserve their vorticity they must lose relative vorticity so that their poleward velocity must decrease (since in these currents relative vorticity $\zeta \sim v_x$). In particular, if the vorticity of such a parcel is βy_0 it cannot progress poleward beyond the latitude of y_0. We conclude that this will be the latitude of the currents required to complete fluid circuits that include the undercurrent.

This vorticity argument may obscure the nonlocalness of the dynamics because it makes no explicit mention of the essential role played by the zonal pressure gradient. In the absence of a zonal pressure gradient, angular momentum conservation would

argue the parcels of fluid arrive at the equator with westward velocity. In fact, the pressure forces accelerate the lower layer flow eastward. Further, for the subsurface flow to have an equatorward component requires a zonal pressure gradient for geostrophic balance. Without this gradient the lower layer fluid demanded by the intense upwelling at the equator is supplied by having the thermocline rise with time. This is what happens in the early stages of the experiment at longitudes where neither the eastward traveling Kelvin nor westward traveling Rossby modes have yet arrived to set up the pressure gradient. Such behavior should be expected from the linear analysis: Eq. (3.1) clearly shows that the layer depth will decrease with time as long as there is no zonal variation. It is only after the initial boundary responses (Kelvin and Rossby) meet that the undercurrent becomes established.

The vorticity arguments presented above may be extended to estimate the latitude y_0 where the undercurrent water originates. Poleward of y_0 the advection of planetary vorticity in the lower layer is balanced by the vortex stretching term fw; y_0 is the point where nonlinear terms enter the vorticity balance. In I Sec. 2 we established that the linear dynamics of the Ekman layer break down at a latitude y_e when inertial terms become important in the Ekman layer. The latitude y_0 must be the same order as y_e since the layers are coupled by vertical motions. These considerations allow us to find a frictional-inertial scaling for Eqs. (1.1), valid when the Ekman layer Rossby number $\epsilon_E \equiv \epsilon\gamma^{-1/2}$ satisfies $\epsilon_E^{-1/3}\gamma^{1/3}<<1$. [See (1.2) for the definitions of ϵ and γ. For the parameter values in Table 1, $\epsilon_E \sim .2$ and $\gamma \sim 10^{-2}$.] The arguments of I give the following rescalings in Eqs. (1.1):

$$y = \epsilon_E^{2/3}\, y'; \quad \alpha = \epsilon_E^{-1/3}\,\gamma^{1/2}\,\alpha'; \quad \gamma_I = \epsilon^{1/3}\,\gamma^{1/2}/\alpha' . \tag{4.1}$$

We assume that the scale in the x direction is long enough so that zonal variations may be neglected except in the pressure gradient term. (It is this term which drives the undercurrent.) It then follows from (4.1) that the remaining variables may be written (cf., Philander, 1971, p. 239):

$$(u^s, v^s, u^1) = \epsilon_E^{-1/3}\,\gamma^{-1/2}\,(u^{s\prime}, v^{s\prime} u^{1\prime}) ; \tag{4.2}$$

$$v^{1\prime} = \epsilon_E^{-2/3}\, v^{1\prime}; \quad w_e = \epsilon_E^{-4/3}\, w'_e ;$$

$$h = \int^x \tau^{(x)}\,(y=0)\,dx + \gamma^{-1/2}\,\epsilon_E^{1/3}\,\rho$$

We now have scalings in terms of the governing parameters for all variables in the equatorial region. In dimensional terms the meridional scale for the equatorial circulation is 1.5°. The scale for the surface velocities and the subsurface zonal velocity is $[\tau_0^4/4\beta\nu_v^2]^{1/3}$; this is about .5 m sec^{-1}. (We have again taken $K \sim 2\nu_v/H$).

Dropping primes and taking $\alpha' = 1$ for simplicity the steady state version of (1.1) is

$$v^s u_y^s + \tfrac{1}{2}\, w_e\,(u^s - u^1) - y\, v^s - \tau^{(x)} = 0 , \tag{4.3a}$$

$$v^2 v_y^2 + \tfrac{1}{2} w_e v^2 + y u^2 \quad \rho_y = 0 , \tag{4.3b}$$

$$v^1 u_y^1 + \tfrac{1}{2} w_e (u^2 - u^1) - y v^1 + \tau^{(x)} = 0 , \tag{4.3c}$$

$$y u^1 + \rho_y = 0 , \tag{4.3d}$$

$$v_y^2 + v_y^1 = 0 , \tag{4.3e}$$

$$w_e - v_y^2 = 0 . \tag{4.3f}$$

The meridional sections of the layer depth that we calculate (Fig. 10) agree qualitatively with the bowing of isotherms which is usually observed beneath the undercurrent (e.g., Knauss, 1966). Eq. (4.3d) requires that the pressure gradient be in geostrophic balance with the lower layer zonal velocity. This accounts for the meridional profile of the pressure gradient. For example, at the center of the basin at .3N the terms fu and $g'h_y$ balance to within 10%. At 3N the balance is within 15%.

c. Longitudinal variations. By considering Eqs. (4.3) at the equator we may easily see that terms neglected in those equations must enter the momentum balance. Since symmetry dictates that $v^1 = v^2 = 0$, (4.3a) says that the vertical advection term balances the wind stress while (4.3c) demands that the same term balance the zonal pressure gradient ($-\tau^{(x)}$). Since the surface stress and the pressure gradient are of opposite sign this is not possible. This argument is not an artifact of our two-level model: it applies to a continuous ocean from the surface down to the undercurrent velocity maximum, a region where w is positive and u_z negative.

It is Eq. (4.3c) that must break down. Two types of terms have been neglected there: those due to vertical friction and those that arise from longitudinal variations. Vertical viscosity is already essential in (4.3) because of its role in the surface frictional Ekman layer; what is at issue here is its direct role in limiting the growth of zonal momentum in the undercurrent. Vertical friction terms enter (4.3) at $O(\epsilon_M^{-1/3}\gamma^{1/3})$ – about an order of magnitude smaller than the retained terms for the parameter values we use. Charney and Spiegel's (1971) model fixes the value of the zonal pressure gradient and excludes other zonal variations *a priori* so that vertical viscous forces offer the only possible balance in a steady state undercurrent. In this context we note that their calculations failed to converge when the eddy viscosity was reduced to 13 $cm^2 s^{-1}$, leaving only the acceleration term to balance the pressure force.

In our model longitudinal variations are permitted and the zonal pressure gradient is free to seek its own value. The usual argument for neglecting longitudinal variations (e.g. Charney and Spiegel, 1971) is that the inertially determined cross stream scale $\epsilon_M^{1/3}$ is so much smaller than the downstream scale L_x. However, it is inappropriate to simply compare length scales because the subsurface velocity components have different scales.

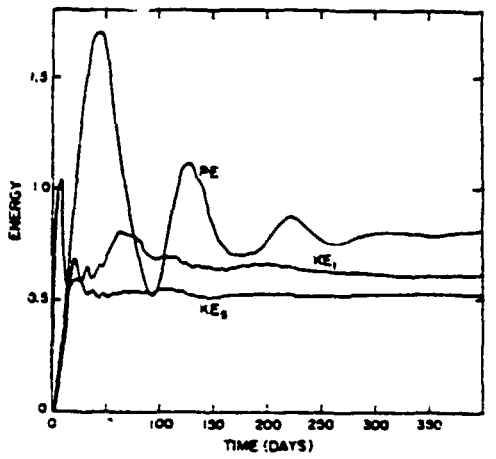


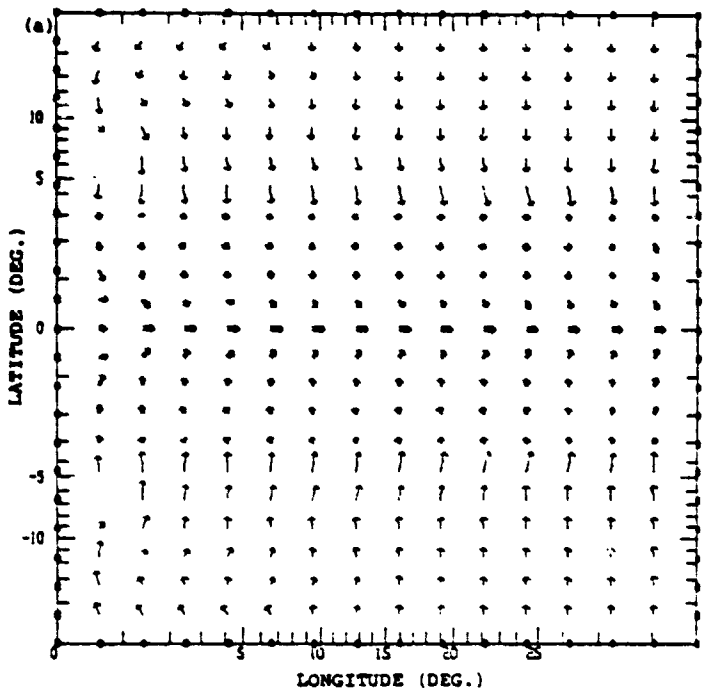
Figure 13. Evolution of energies in the equatorial region for the nonlinear response to a westerly wind. Like Fig. 3.

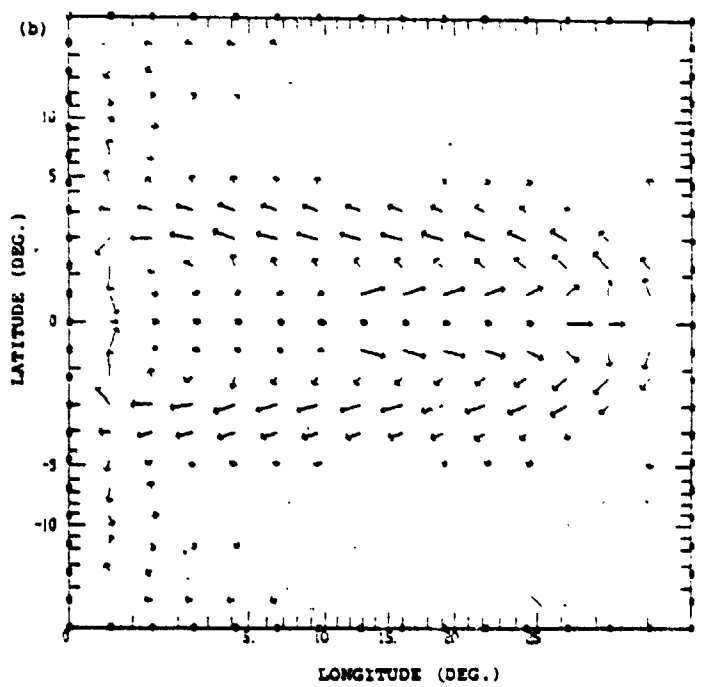
In (4.3c) terms involving zonal variation (i.e., $u^1 u_x^1$ and p_x) will enter before the vertical viscous terms as long as the zonal length scale L_x is less than $\gamma^{-1}\,\epsilon_B^{2/3} = 10^4$ km (for our parameter values). Hence these terms are more important than the frictional ones. The most stringent condition for neglecting zonal variations altogether in (4.3) is that $L_x >> \gamma^{-1/2}\epsilon_B = 1300$ km. This is large enough so that we should expect zonal variations to be dynamically significant, an expectation borne out by our numerical experiment. [It is the downstream acceleration $u^1 u_x^1$ that balances the pressure gradient in (4.3c).]

5. Nonlinear response to a uniform westerly wind

We now consider the nonlinear response to a uniform westerly wind stress. The only parameter changed from the previous section is the wind direction. The responses in the two *nonlinear* cases are quite different because the beta effect causes east-west asymmetries in the ocean dynamics. On the other hand, the *linear* responses are very much the same. The east wind response (Sec. 3) becomes the west wind response simply by changing the sign of all variables (h into $-h$, u^a into $-u^a$, etc.); the *pattern* of the response is unaltered.

The energy plot for this case (Fig. 13) resembles that for the linear east wind (Fig. 3) more than the nonlinear one (Fig. 8). The potential energy curve for the present case is almost identical to that for the linear case: the amplitudes are approximately the same and the oscillations have the same periods. In Sec. 3 these oscillations were shown to result from the reflections of Rossby and Kelvin waves at the meridional boundaries and the same explanation applies here. Unlike the east wind nonlinear case, the undercurrent does not become strong enough to significantly


(a)
LATITUDE (DEG.)
LONGITUDE (DEG.)


(b)
LATITUDE (DEG.)
LONGITUDE (DEG.)

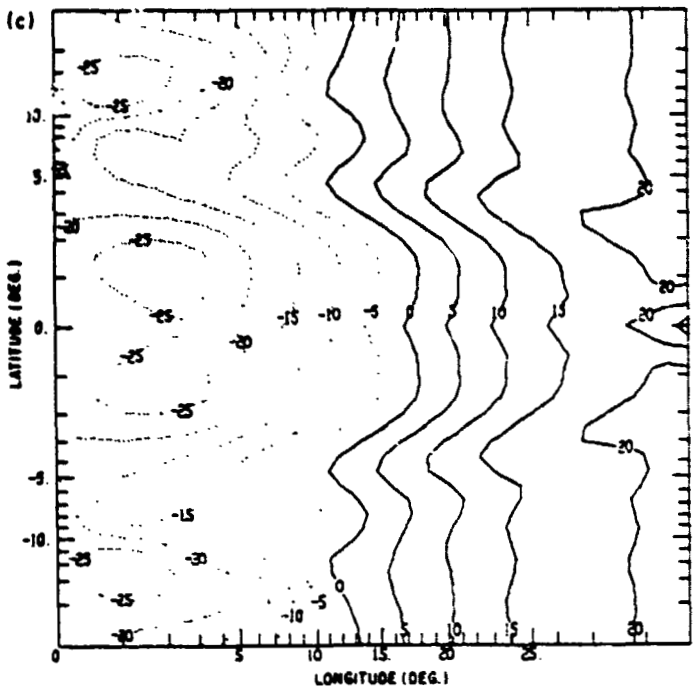


Figure 14. Nonlinear response to a westerly wind at day 398. Like Fig. 7 (a) u' (b) u² (c) *h*. Contour interval is 5 m.

interefere with the Rossby wave reflections at the eastern side. The layer depth evolves much as it did in the linear case. In contrast to the east wind nonlinear case (and like the linear one) the final zonal pressure gradient very nearly balances the wind stress. The inclusion of vertical advection of momentum makes the distribution of kinetic energy different from the linear case, with the surface layer being less, and the lower layer more, energetic. In this it is like the other nonlinear case but the lower layer kinetic energy is much smaller here.

At the equator the vertically integrated zonal transport $\tilde{u}$ becomes close to its final interior value within 8 days though there is some oscillation about this final value (of about 75 $m^2\ sec^{-1}$) until about day 160. The nonlinear east wind case takes about 24 days to reach a value of 75 $m^2\ sec^{-1}$ and about 80 days to approximate its final value of 140 $m^2\ sec^{-1}$. In both the east and west wind cases the steady state requires westward currents off the equator to return the water that has travelled to the eastern side at the equator. In the west wind case these lower layer currents, centered at 2S and 2N, take longer to become established. They are absent before day 40 and take 80 days to approach their final value, so the early evolution supplies the net eastward transport of water needed to set up a zonal pressure gradient opposite to the wind. When the wind is from the east the westward flowing currents are set up before the eastward flowing undercurrent.

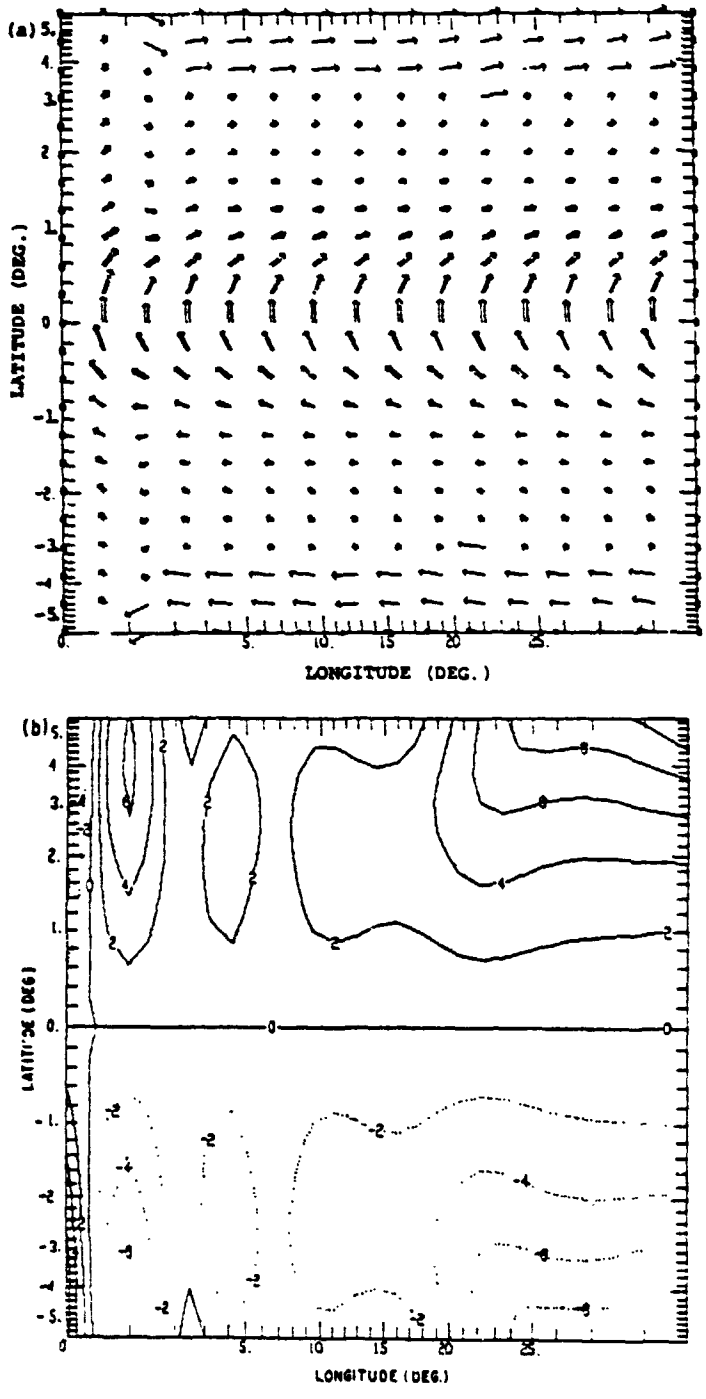


Figure 15. Linear response to a southerly wind at day 40. Like Fig. 6 but only the equatorial region (5.6S to 5.6N) is shown. (a) u' (b) h. Contour interval is 2 m.

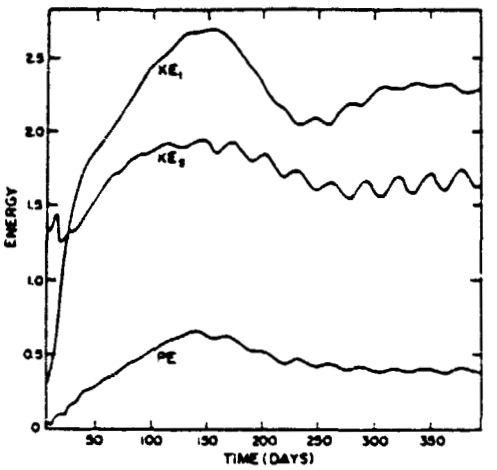


Figure 16. Evolution of energies in equatorial region for the nonlinear response to a southerly wind. Like Fig. 3. Initial time is 5 days.

ment proceeding from east to west and from the equator poleward. No Kelvin waves are excited so the adjustment is effected entirely by long wave Rossby modes. Because of the absence of Kelvin waves there is nothing like the oscillation shown by the linear response to an east wind. Since bottom friction is present the steady state has an equatorial boundary layer with nonzero transports. The characteristics of this layer are in detailed agreement with the analytic results of I. By far the largest part of the ocean's energy in the equatorial region is in the surface layer motions. As in the previous linear case there is little kinetic energy in the lower layer. Since $h \lesssim y_T^{(x)}$ for all time there is little potential energy near the equator where y is small.

Figure 15b shows the layer depth at 40 days. The northward tilt is clearly evident at the eastern side. The lower layer currents are in geostrophic balance (except for the area along the equator where the pressure force is frictionally balanced by southward flow). At the western side h has strong gradients to balance the meridional boundary currents evident in both layers. The southward current east of the boundary current is required by conservation of potential vorticity. Mathematically, it results from the Bessel function behavior of the western boundary layer solution (cf. CSII, Fig. 2).

7. Nonlinear response to a uniform southerly wind

An overview of the spin up process is given by Figure 16, which depicts the energy in the equatorial region. As in the linear case, the surface layer kinetic energy quickly (order 8 days) rises to within e^{-1} of its final value as the wind stress transfers energy to the ocean. Thereafter, the increase in surface energy and potential energy continue until about day 150. The lower layer kinetic energy becomes

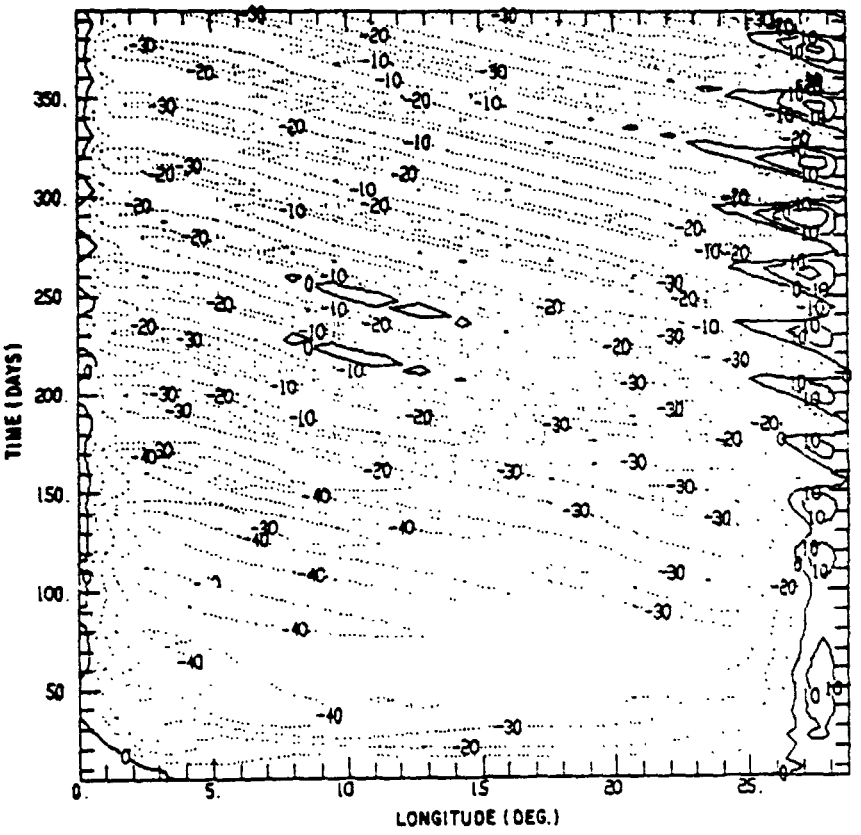
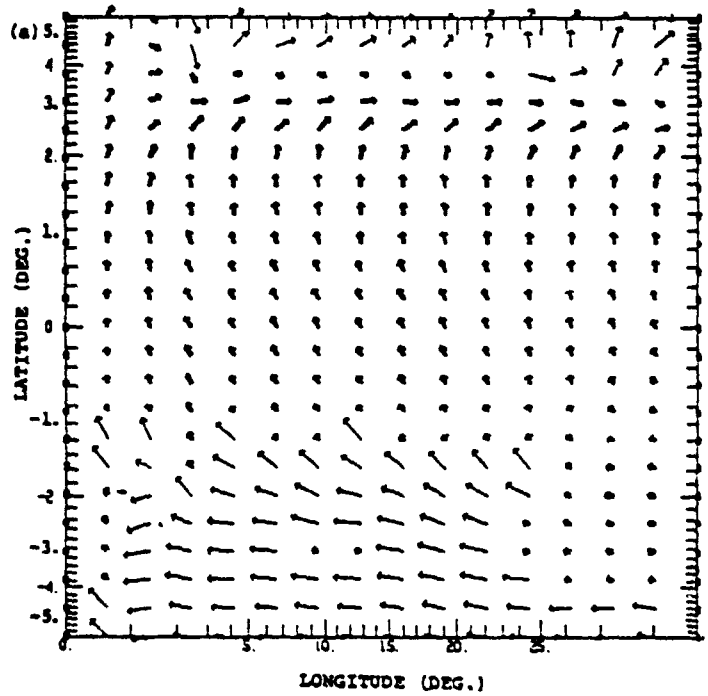
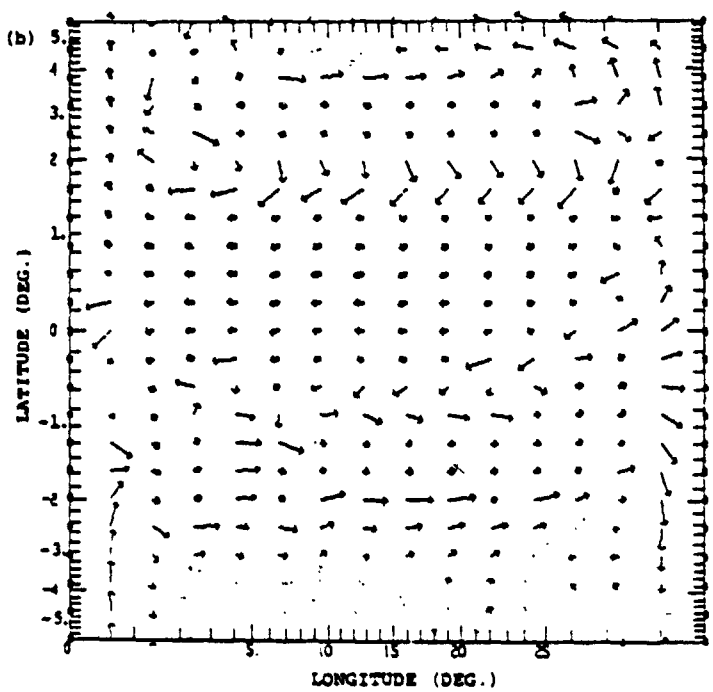


Figure 17. Contours of lower layer zonal velocity, u^1, along the equator as a function of time for the nonlinear response to a southerly wind. Contour interval is 10 cm sec^{-1}. Note the regular westward phase progression.

almost an order of magnitude larger than in the linear case indicating the importance of vertical advection as a mechanism for transferring momentum to the lower layer.

Beginning at about day 100 an oscillation with a period of 29 days may be observed in the surface kinetic energy. At about 150 days, an oscillation in the potential energy sets in approximately in phase with this. The lower layer kinetic energy starts to decrease, eventually leveling off to oscillate about a steady value, the oscillations being out of phase with those in the other quantities. This suggests an instability which draws its energy primarily from the kinetic energy of the flow. Figure 17 shows a plot of phase lines of the lower layer zonal velocity at the equator—the abscissa is distance along the equator, the ordinate is time. Beginning near the western side at about day 100 and appearing later at the eastern side, a very regular progression of phase from east to west may be observed. (Similar plots of the other variables give essentially the same picture.) These waves have a period of 29 days and a phase speed of 32.5 km/day, giving a wavelength of 950 km.


(a)
LATITUDE (DEG.)
LONGITUDE (DEG.)


(b)
LATITUDE (DEG.)
LONGITUDE (DEG.)

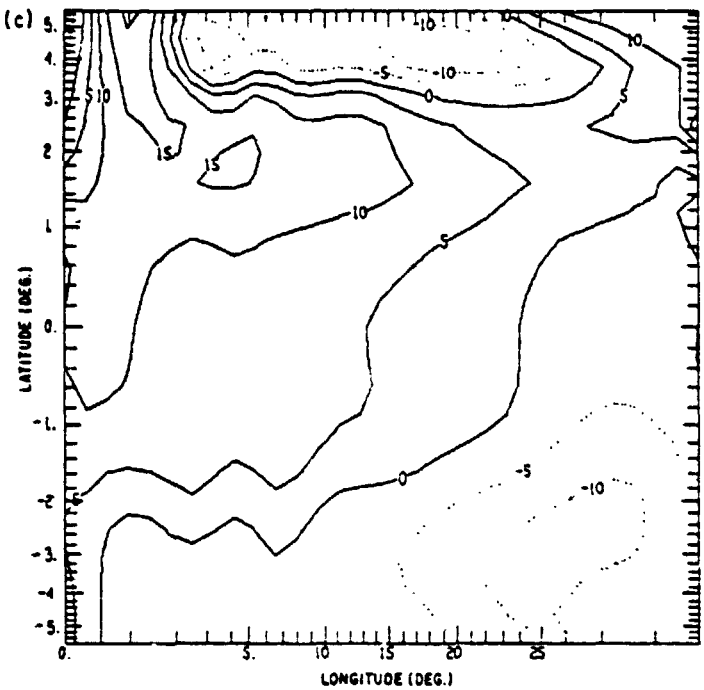


Figure 18. Nonlinear response to a southerly wind at day 40. Like Fig. 15. (a) u' (b) u^1 (c) h. Contour interval is 5 m.

a. Evolution of the flow. In the early stages many of the features of the flow pattern can be understood by considering the kinematic effects of the vertical and meridional advections on the linear response. At 8 days the circulation pattern in both layers is similar to the linear response at the same time, but some departures from the linear symmetries are already apparent. In the surface layer the maximum meridional velocity occurs at approximately 1N, rather than on the equator as in the linear response. This difference is due to the self-advection of northward momentum by the surface currents near the equator. A similar advective effect is observable in the zonal component of the surface current. Westward momentum has been advected northward across the equator so that the surface flow is westward to 1.2N. The eastward momentum in the surface layer north of the equator has also been advected northward—but not beyond 3N, where the meridional velocity goes to zero. The effect is to compress the eastward flow into a narrower, more intense jet. The eastward flow at 2.5N is at speeds of .9 ms^{-1}, compared to the linear maximum of .6 ms^{-1} at 1.2N. The trend is thus toward the development of an eastward jet at 2.5N with a broader, slower westward flow at the equator. At this time, vertical velocities are negligible everywhere in the interior and currents in the lower layer are small.

The results of this development are evident at day 40, Figure 18 (cf. the linear case Fig. 15). The eastward jet is now centered at 3N. There is a considerable horizontal convergence into the jet, resulting in substantial downwelling; elsewhere in the interior the vertical velocity is negligible. This downwelling advects eastward momentum into the lower layer so that the flow there is also eastward (Fig. 18b). The result is a large vertically integrated transport to the east—a factor of 5 larger than in the linear case.

b. The x, t independent flow. The features of the flow evident at day 40 were substantially established within 16 days. The pattern persists with little qualitative change beyond day 100. We note that the interior flow is approximately steady and independent of longitude. [The primary exception to this is the layer depth (Fig. 18c) which shows a more uniform tilt to the north at the eastern side, similar to the linear case. It also shows a suggestion of a wavelength structure at about 4N.] The part of the interior field that is independent of x and t has four dynamically distinct regions:

(i) South of about 2.5S the response is essentially linear, like that discussed in the previous section. The dominant feature is the surface wind drift current to the left of the wind.

(ii) From 2.5S to about 2.5N the surface flow turns from northwestward to northeastward to due east. Vertical velocities are everywhere upward and small [$0(10^{-5}$ m sec$^{-1})$], with most of the upwelling south of the equator. South of 1N u^1 is westward, while north of 1N it is eastward, with a magnitude comparable to u^0 near the equator. The meridional component is southward everywhere.

(iii) From 2.5N to 5N there is an eastward zonal jet in both layers: at 3N u^0 is as high as 1.2 m sec^{-1}; u^1 is over .4 m sec^{-1}. There is strong surface convergence into the jet with large downwelling at its core ($w = 3 \times 10^{-4}$ m sec^{-1}).

(iv) North of about 5N the model response again becomes wind drift dominated and essentially linear.

This description is in close agreement with the x-independent, steady state calculation of Charney and Spiegel (1971). (See their Figs. 11 and 12.) The only notable disagreements are that their surface velocity in the jet is smaller (less than 1 m sec^{-1}), their downwelling region is broader, and their upwelling region narrower than ours. We now seek a simple model (independent of x and t) to elucidate the physics of this flow.

Regions (i) and (iv) are explicable in terms of linear dynamics. Now consider the surface flow in region (ii). A parcel in the vicinity of the equator will acquire a northward velocity component (frictional forces give it a component in the direction of the wind). As it moves northward, it acquires cyclonic planetary vorticity. Since it approximately conserves its total vorticity, it must acquire anti-cyclonic

relative vorticity. The effect is to turn the parcel clockwise toward the east. As long as the parcel moves northward, it is able to acquire energy from the wind stress. At some latitude the parcel's northward momentum is being converted into eastward momentum more rapidly than it is replenished by the wind. Eventually, the parcel will be travelling due east, still carrying the approximately zero total vorticity it had near the equator. To the north of this, the flow is in the wind drift regime where the vorticity of surface parcels is approximately the local planetary vorticity. The transition between the two flow regimes demands a shear layer in which the surface eastward velocity is reduced to the north, thus adding enough positive vorticity to the flow to match it to the planetary vorticity. This accomplished by the downwelling in the jet which transports the eastward momentum downwards.

We may formulate a simple model for the surface flow in region (ii) in order to obtain some quantitative checks to accompany this qualitative description. Neglecting friction the approximate governing equations are

$$\frac{d}{dt} u^e - \beta y v^e = 0 , \tag{7.1}$$

$$\frac{d}{dt} v^e + \tfrac{1}{2} \frac{w}{\eta} + \beta y u^e = \frac{\tau}{\eta} \tag{7.2}$$

Since the flow varies little with t or x, $d/dt \equiv v\partial/\partial y$ and $w/\eta \equiv v_y{}^e$. Since $v^e u_y{}^e >> v_y{}^e v^e$ the vertical advection term is neglected in (7.1).

An energy equation (which allows for some loss to the lower layer) may be formed from (7.1) and (7.2):

$$\tfrac{1}{2} \frac{d}{dt} \left(u^{e2} + \frac{3}{2} v^{e2} \right) = \frac{\tau v^e}{\eta} .$$

By making use of the definition $dy/dt \equiv v$, this may be integrated to yield

$$u^{e2} + \frac{3}{2} v^{e2} = 2 \frac{\tau}{\eta} y + \text{const.}$$

Eq. (7.1), which expresses the conservation of vorticity, integrates to

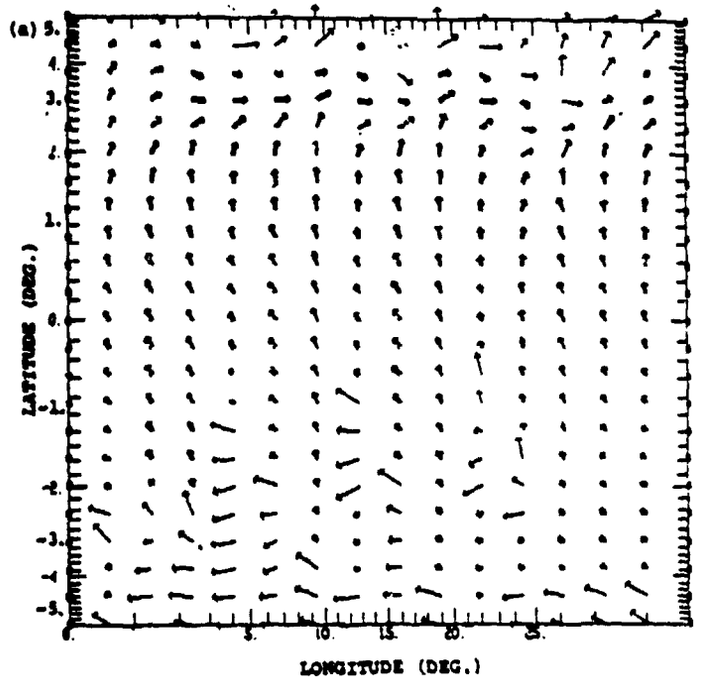
$$u^e (y) = \tfrac{1}{2} \beta y^2 + u^e (0) .$$

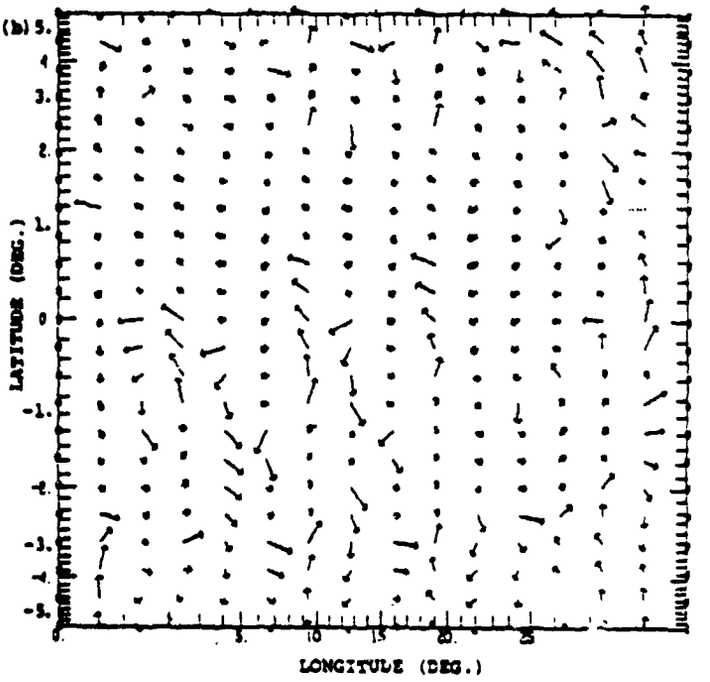
We may now simplify things further by assuming that $u^e = v^e = 0$ at the equator and so obtain

$$u^e \sim \frac{1}{2} \beta y^2 ; \quad v^{e2} = \frac{4}{3} \frac{\tau}{\eta} y - \frac{\beta^2 y^4}{6} . \tag{7.3}$$

The position of the jet is at latitude y_j where $v^e \sim 0$:

$$y_j \sim 2 \left[\frac{\tau}{\eta \beta^2} \right]^{\frac{1}{3}} \sim 340 \text{ km} \sim 3^\circ .$$


(a)
LATITUDE (DEG.)
LONGITUDE (DEG.)


(b)
LATITUDE (DEG.)
LONGITUDE (DEG.)

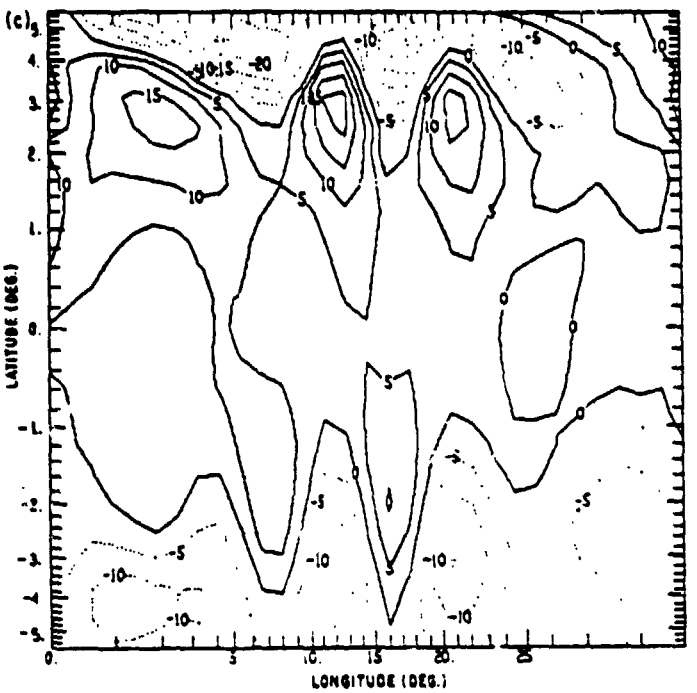


Figure 19. Nonlinear response to a southerly wind at day 398. Like Fig. 18 (a) u' (b) u^1 (c) h.

At this latitude $u_J = \frac{1}{2} \beta y_J^2 \cong 1.2$ m sec^{-1}. These values are in excellent agreement with the numerical calculation. We may also obtain a scale for the other velocity component by considering the latitude y_m where v^0 is a maximum. We obtain $y_m \sim 2°$ and $v_m \equiv v_s(y_m) \sim .7$ m sec^{-1}, again in good agreement with the numerical calculation.

Turning now to the jet itself, the requirement that the vorticity of the flow be brought up to the local planetary vorticity in order to match onto the linear regime gives a scale for the width Y of the shear zone:

$$\beta (Y + y_J) \sim u_J/Y \rightarrow Y \sim 1.1°$$

This is the right order but slightly too wide (the model results show $Y = .8°$). One feature of this description which agrees well with the numerical calculation and that of Charney and Spiegel (loc. cit.) is that the zonal velocity of the jet falls off more rapidly to the north than to the south.

The fluid which descends in the jet arrives in the lower layer with considerable eastward momentum and negative relative vorticity. In region (ii) continuity requires that v^1 be southward. Parcels will approximately conserve their total vorticity because vertical exchanges are small and because both f and variations in the layer depth are small (so that the variation of potential vorticity is given by the variation

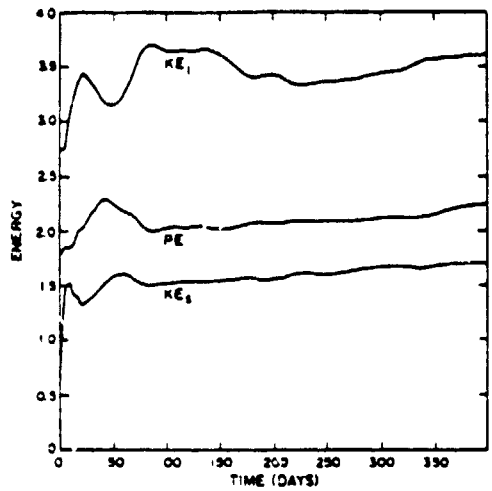


Figure 20. Evolution of energies in the equatorial region for the response to a southeasterly wind. (At $t = 0$ the flow is adjusted to an easterly wind. A southerly component is added so that the wind is then from the southeast.) Like Fig. 3.

of vorticity). As a parcel travels southward, its planetary vorticity decreases so its relative vorticity ζ must become less negative and may even become positive. In our calculation $u_y^1 = 0$ at about .3N. North of this point the flow impinging on the western boundary turns clockwise to the north ($\zeta < 0$), while south of it the flow turns counterclockwise to the south ($\zeta > 0$).

c. Stability. The foregoing analysis has implications for the susceptibility of the flow to shear instability. For nondivergent inviscid flow a necessary (but not sufficient) condition for instability is that the vorticity—$f - u_y$ in our case—have an extremum. Though our situation is more complicated, this simple criterion will serves as a useful guide (see below). This condition is usually not met by geophysical flows because the gradient of planetary vorticity, $\beta - u_{yy}$, is always positive. In region (ii) the flow in each layer is characterized by the conservation of total vorticity, thereby neutralizing the stabilizing effect of beta. The flow is thus marginally unstable; and, in fact, the model calculation does exhibit numerous extrema in the profile of $f - u_y$.

At day 120, a time when the energy curves (Fig. 16) show some evidence of instability, the waviness suggested at day 40 (Fig. 18) at 4N is marked, with a similar wavy pattern developing at 4S as well. At 160 days, when the lower layer kinetic energy reaches a peak, this pattern is no longer confined to the western side, but is present in equal amplitude across the width of the basin. By about day 200, the flow settles into a repeated pattern with wavelike disturbances propagating (in the phase sense) across the basin from east to west (see Fig. 17), though the zonally averaged mean flow is still similar to that at day 40. Figure 19 shows the fields during this final period of the flow's evolution. The variations in the layer depth (Fig.

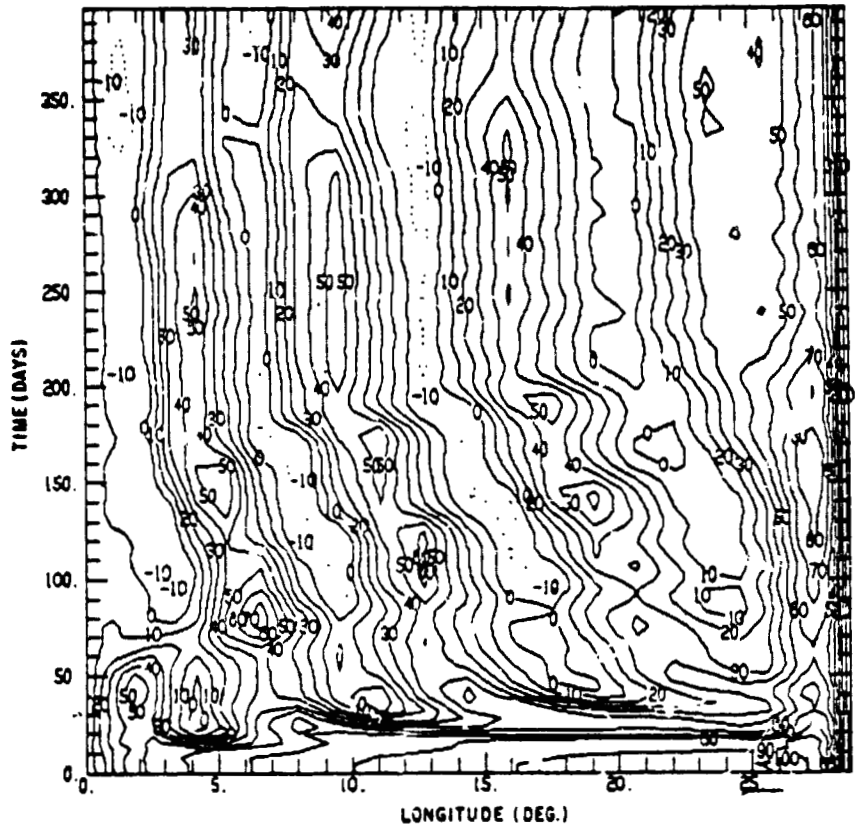


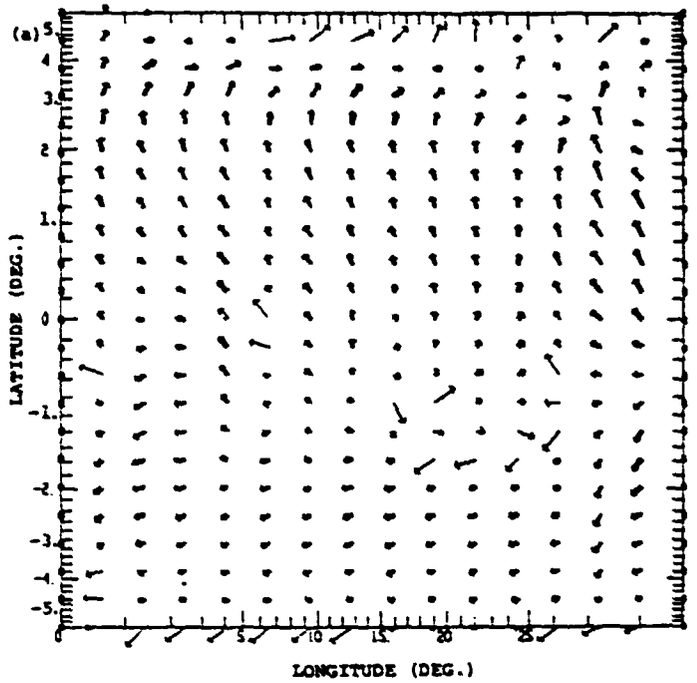
Figure 21. Contours of u' along the equator for the nonlinear response to a southeasterly wind. Like Fig. 17.

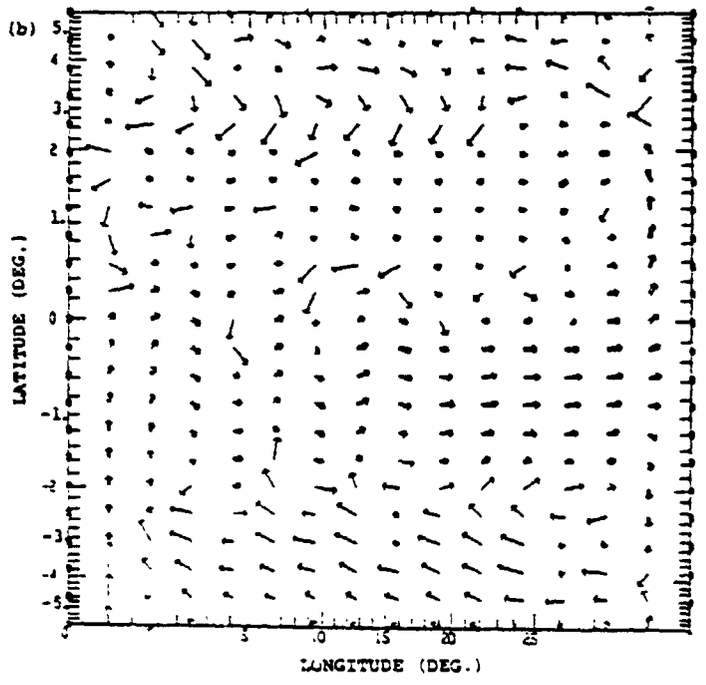
19c) are dominated by this instability from 7S to 7N, while poleward of these latitudes it exhibits a general south to north tilt similar to the linear case.

It is of interest to compare our results with the results for the stability of equatorial currents given by Philander (1976). On the basis of the vertically averaged zonal velocity we can crudely fit the model currents to a sech^2 profile, viz. $U \cong U_0 \, \mathrm{sech}^2 \, y/L + U_1$, by taking $U_0 \cong -.7$ ms^{-1}, $U_1 = +.4$ ms^{-1} and $L \cong 100 - 200$ km. Then $R_i = gHU_0^{-2} \cong 8$ and $R_0 = U_0 \, (\beta L^2)^{-1} \sim -1$ to -4. From Figure 3 of Philander (loc. cit.) the wavelength of the fastest growing wave for these parameter values is approximately $2\pi L$—between 600 and 1,200 km in our case. This is consistent with the model results.

8. Nonlinear response to a uniform southeasterly wind

In this case the initial state is taken to be the steady state response to a uniform easterly wind described in Sec. 4. (To be precise, the initial state is taken as the





C-2

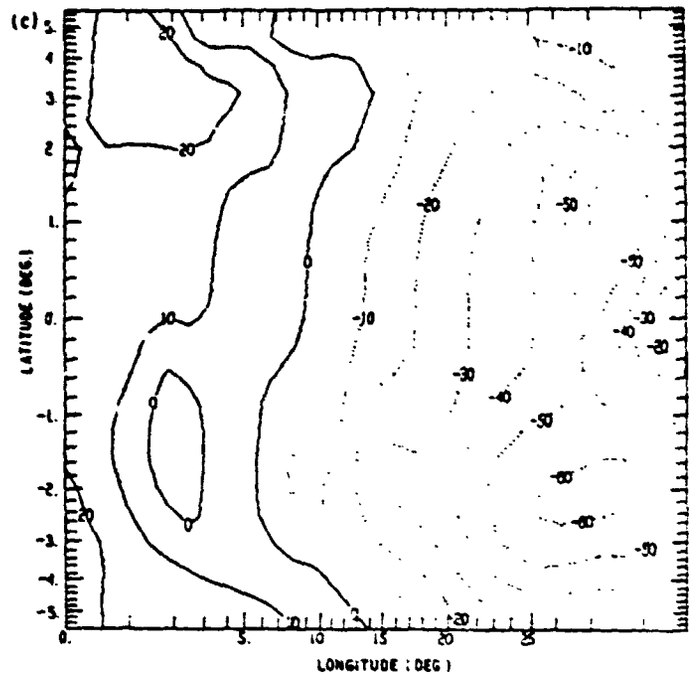


Figure 22. Nonlinear response to a southeasterly wind at day 40. Like Fig. 18 (a) u^* (b) u^1 (c) h. Contour interval is 10 m.

state which resulted after 384 days; it is barely distinguishable from the one at day 398 depicted in Fig. 12.) At $t = 0$ a southerly wind component is added to the prevailing easterly component so that each wind component is .465 dynes cm^{-2}. This is intended to be a very crude analog of the southeast monsoon that occurs in the Atlantic in the late Spring. The linear response to a southeast wind is simply a superposition of the linear responses to a south wind and to an east wind but the nonlinear response is more complex.

The evolution of model energies, Figure 20, shows that after 400 days the model ocean still has not reached a truly steady state, even in the equatorial region. (The continued increase in PE and KE_1 appears to be due to a strengthening of the current at 4.5N. See Figure 22.) After day 80 the potential and upper layer kinetic energy do not vary greatly. The lower layer kinetic energy takes about twice as long to become approximately constant. In the equatorial region only the upper layer kinetic energy is substantially different from its value at $t = 0$.

Figure 21 is a plot of contours of u^1 at the equator with time as the ordinate. It is similar to Figure 17, which showed a regular progression of phase for the wave-like instability that arose in that south wind case. In the present case, in the time period from about day 25 to about day 175 there is some *apparent* phase propagation to

the west. However, none of the lines of constant amplitude cross the basin, and after day 175 all such east to west movement ceases. The pattern of evolution resembles the way in which the mixed mode in the linear south wind case contracts toward the western boundary with time (due to its Bessel function behavior; cf. CSII). We will return to this point after considering the early evolution of the flow.

The two most prominent developments in the first 40 days (see Fig. 22) are the southward shift of the undercurrent and the development of an eastward current between 3N and 5N. The latter resembles the south wind response, Sec. 7. The southward shifting of the undercurrent is already evident at 8 days. At the same time the surface currents near the equator have become westward. This upwind shifting of the undercurrent in the presence of meridional winds has been found in earlier theoretical investigations (Robinson, 1966; Charney and Spiegel, 1971) and has been observed in the world's oceans (e.g., Taft and Knauss, 1967).

The surface flow (Fig. 22a) strongly resembles the south wind response (Fig. 18a). The east wind influence shows most clearly in the poleward wind drift currents south of 1S. The lower layer currents (Fig. 22b) show effects from both the south wind (cf., Fig. 18b) and the east wind (cf., Fig. 12b displaced south). The region of eastward flow centered at about 1S is broader than is the case with either of the simpler wind systems. The eastward currents at 4N induced by the southerly wind component appeared first at the western side of the basin and gradually extended to the east. The layer depth still resembles its initial state (Fig. 12c) far more than the south wind response (Fig. 18c).

Between day 8 and day 40 the most important feature to appear in the lower layer is the wavy pattern which causes the undercurrent to meander about its mean latitude of about 0.6S. These meanders persist thereafter though their form changes. The disturbance propagates from west to east in the sense that it appears earlier at the western side. Figure 21 showed that any phase propagation is westward, but to speak of phase propagation is misleading. The meander pattern migrates westward over the first 175 days and then remains stationary. The result might be described as a standing wave of zero frequency. The structure of these meanders is most clearly revealed by the contours of the zonal and meridional transports at day 398 shown in Figure 23. They have their largest amplitude between 0 and 2S; with the amplitude decreasing from west to east. The wavelength of the meanders is about 650 km and shows a slight increase from west to east. We offer the interpretation that these meanders are due to a mode generated at the western boundary in response to the south wind. Such a mode is the nonlinear analog of the mixed mode that is generated when the initial state is a resting one. It plays the role of a barotropic instability in the sense of acting to reduce the horizontal shear of the zonal currents.

In addition to these meanders, the other prominent features of the flow in the equatorial region at day 398 are the eastward jet in both layers centered at 4N and

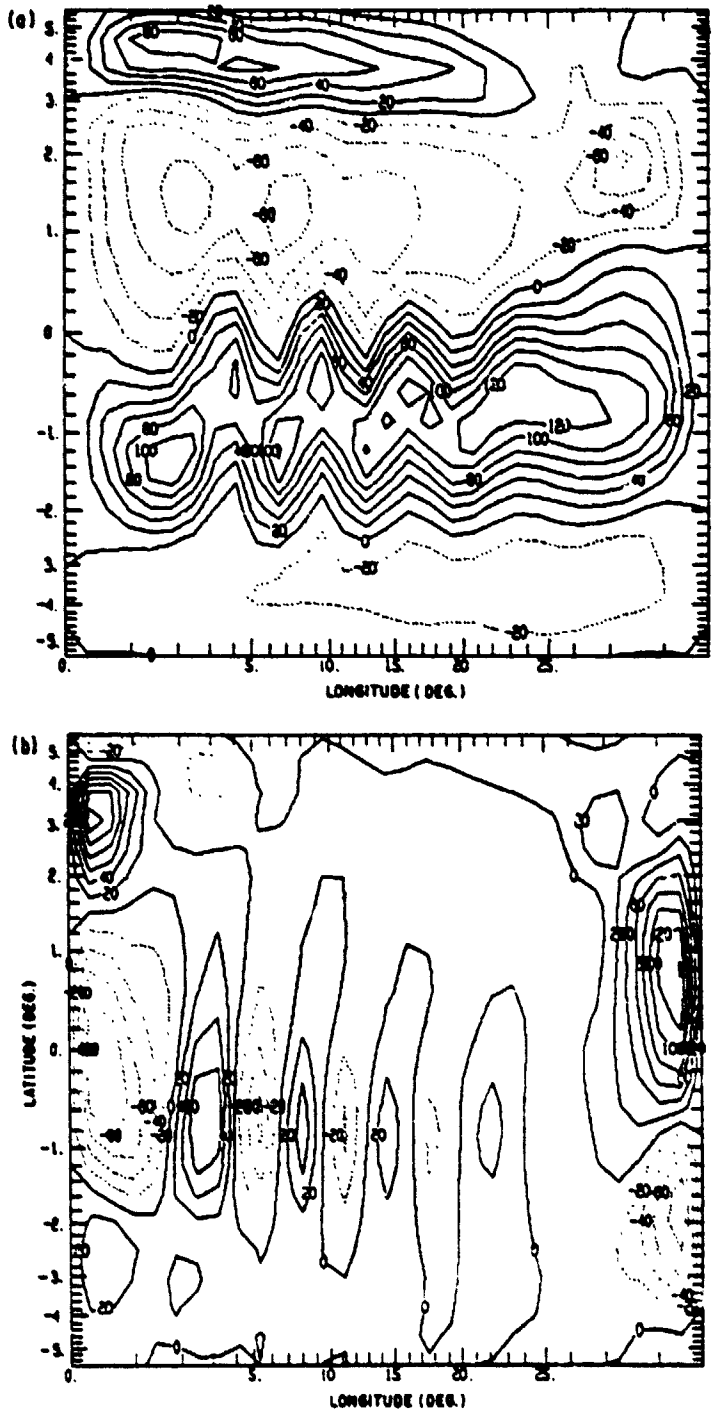


Figure 23. Contours of the vertically integrated transports in the nonlinear response to a southeasterly wind at day 398. Contour interval is 20 m^2 sec^{-1} (a) Zonal transport, ϕ. (b) Meridional transport, ϑ.

the undercurrent with a mean position at about 0.7S. Elsewhere between ±5° the lower layer currents are generally westward. Except that the jet is slightly further north and its total transport less than in the south wind case, the mean flow between the equator and about 5N is very similar in the two cases. The dynamical arguments presented in Sec. 7 apply here as well. South of the equator the two cases are quite different due to the presence of the undercurrent at about 1S and an additional region of westward transport to the south of it. The flow pattern between 3S and 3N that we find at day 398 resembles the similar calculation made by Charney and Spiegel (1971). (See their Figs. 9 and 10; they only show the region from 3S to 3N.) Flow in the meridional plane is similar in the two calculations, with the division between northward and southward surface flow occurring at the latitude of the undercurrent. Of course, since their model has no zonal variation, the meanders are absent.

9. Discussion

Timescales. The analysis in I showed that there are two timescales governing the flow. The first of these is the time for the transfer of wind energy to the upper frictional (Ekman) layer; this is 0 (20 days) in all cases. The second is the time for the sea surface to set up and establish the basin-wide pressure gradients. Away from the equator this setup time is that given by the linear analysis because the extra-equatorial dynamics are linear. (From our viewpoint the nonlinearities of the western boundary current region are a local effect that does not influence the large scale adjustment.) The setup time [0(1 year) at 10°] decreases toward low latitudes with the equatorial time varying greatly from case to case. Nonlinear effects may either lengthen or shorten the corresponding linear response time, with the result that the nonlinear time varies between 150 and 300 days. This setup time varies linearly with the longitudinal extent of the basin (for zonal winds, see below). The overall setup time for each of the world's equatorial oceans is thus longer than or comparable to the characteristic time for the major wind stress variations associated with the seasonal cycle. The oceans do not have enough time to adjust to be in equilibrium with the winds (eg. Sverdrup balance).

On the other hand, important features of the equatorial circulation are qualitatively present quite early. In all the zonal wind cases there is a significant zonal pressure gradient everywhere on the equator at 15 days, the time when the effects of the meridional boundaries are first felt at all longitudes. This occurs when the initial Kelvin and $n = 1$ Rossby modes meet. Nondimensionally this time is $3/4T$ where T is the time for a Kelvin wave to cross the basin. For the Indian and Atlantic Oceans the time T is approximately one month while for the Pacific it is three months, so boundary responses to seasonal variations have time to extend across each ocean. Another important milestone occurs at a time $4T$ (80 days in the experi-

ments) when the first reflections (i.e., the $n = 1$ Rossby reflection of the initial Kelvin mode and the Kelvin reflection of the initial $n = 1$ Rossby mode) cross the basin. By this time enough of the zonal rearrangement of mass has been accomplished so that the circulation qualitatively resembles the steady state. If anything, the time $4T$ is more significant in the nonlinear cases because the nonlinearities tend to damp subsequent oscillations, especially with easterly winds. To the extent that the flow at this time is in equilibrium with the winds the Atlantic and Indian Oceans will be in equilibrium with the annual component of the winds and the Pacific will not.

Nonlinearity and waves. In all cases it is the surface flow on (and near) the equator that first becomes nonlinear; this occurs within three or four days (using a local Rossby number as a measure). Within two weeks nonlinear distortions of the flow field are evident in both layers. Vertical velocities are large and vertical advections play an important role in the lower layer circulation.

While the response to zonal winds involves both wave dynamics and nonlinear currents, the southerly wind response (with both resting and east wind initial states) is fundamentally nonlinear and wave concepts are inapplicable. It bears little resemblance to its linear counterpart and the setup time is unrelated to that given by linear theory. Adjustment occurs primarily in the meridional plane so the setup time is largely independent of the longitudinal extent of the ocean. Our results show that in the first month the ocean rapidly approaches its equilibrium value, changing only gradually thereafter (at least until the onset of instability). Since the real oceans are not initially at rest their response times to seasonal variations may differ somewhat from those given above. However, the setup time for a southerly wind was about the same for an initial state with an undercurrent and one at rest.

Undercurrent dynamics: easterlies and westerlies. We have seen that the model's response to easterlies (the prevailing winds in the equatorial Atlantic and Pacific) is rather different than its response to westerlies (as are present in the Indian Ocean with the onset of the southwest monsoon; see Wyrtki, 1973 and Knox, 1976). In the case of easterlies the undercurrent requires an eastward pressure gradient, not only as a direct driving, but also to produce equatorward flow. This undercurrent is nonlocal in character. It becomes established at a given longitude only when the meridional boundaries are felt there.

By contrast, the eastward flow at depth in response to westerlies is locally determined. The dynamics are frictional and inertial, relying upon frictionally induced surface convergence and associated downwelling to produce eastward momentum at depth. Because the response is local it is also quite rapid: within a week the undercurrent is nearly at full strength. The data collected by Knox (1976) at Gan is strong evidence for the rapid generation of an undercurrent after the onset of westerlies.

These observations also show that the currents reverse when the winds slacken or become meridional. Knox attributed this to westward flow driven by the pile up of

water at the eastern side, as in O'Brien and Hurlburt's (1974) numerical calculation. The same feature is present in our *linear* zonal wind response (cf. Sec. 3 and Fig 4) and was attributed to the eastern boundary Rossby waves overbalancing the wind stress. However, the transit time from the eastern boundary to Gan for such waves is not consistent with the observations [at least for the first baroclinic mode; but cf. Philander (1978b)]. Such a current reversal is not present in our *nonlinear* calculation where the pressure gradient slows the inertial eastward flow but is not strong enough to reverse it. Our model results suggest another explanation. Since the subsurface eastward flow is driven by local winds, the cessation of the wind eliminates the source of eastward momentum and the undercurrent can disappear as rapidly as it appeared. The basin-wide westward pressure gradient previously set up by the westerlies led to subsurface poleward flow which can now act to carry eastward momentum away from the equator. At the same time right on the equator it will drive a down gradient flow to the west. In addition, the calculation of Sec. 7 suggests that flow at the equator will be westward in the presence of meridional winds. (This is true whether the winds are northerly or southerly). Knox's observations are not inconsistent with this, but since the dynamics involve the equatorward advections of both surface and subsurface currents some knowledge of the currents off the equator is needed to decide whether this applies.

Downstream variations in the undercurrent. Our undercurrent calculation resembles the observed undercurrent in many important respects. It also shares many features with the x-independent model of Charney and Spiegel (1971), but permitting zonal variations and not constraining the pressure force to balance the zonal wind stress makes for some important differences. The pressure gradient was larger in our model, making it more in line with observational evidence (Montgomery and Palmen, 1940; also see Charney, 1960, p. 305). The terms uu_x and vu_y are of comparable magnitude. Previous theoretical studies of the undercurrent in homogeneous oceans have neglected the former term by arguing that the zonal length scale is much greater than the meridional one. However, in the undercurrent $u>>v$, so this argument breaks down. We have shown that either zonal variations or friction terms must enter the undercurrent momentum balance. For reasonable values of the coefficients of eddy viscosity and realistic basin sizes, zonal variations will enter before additional friction terms. (Observational accounts of the momentum balance in the undercurrent [Knauss, 1966; Taft, *et al.*, 1974) have also neglected downstream advections, but this is due primarily to a lack of data.]

The model results show an increase in transport downstream. Such behavior is observed in the western Pacific where the depth of the thermocline shows little variation (as is the case with the model's thermocline). In the eastern Atlantic and Pacific, where the thermocline becomes very shallow, undercurrent transports appear to decrease downstream. To reproduce this feature it appears to be necessary to include

some physical mechanism which allows the pressure gradient force to be uncoupled from the vertical extent of the undercurrent. One such possibility is the equatorial effect of the thermohaline circulation (Philander, 1973), though it may be enough for the model to allow downwelling beneath the undercurrent core. Observational evidence does not enable one to determine with certainty if the loss of fluid from the undercurrent occurs in the meridional or the vertical plane.

Countercurrent dynamics. Sverdrup's (1947) explanation of the relationship between the North Equatorial Countercurrent and the curl of the wind stress is one of the great triumphs of dynamical oceanography. Studies based on more recent observational data (Kendall, 1970) have shown some discrepancies from the Sverdrup balance. The model results in the southerly and southeasterly wind cases suggest a possible cause for these discrepancies. In both of these cases a strong eastward jet developed in both model layers as a response to the southerly wind component. In the pure south wind case this jet is centered at 3N; with southeasterlies it is at 4N. In both cases it extends to 5N and so would overlap the wind stress curl driven countercurrent at its southern edge The evolution of this countercurrent was explained in terms of the kinematic effects of advection on the linear response in the frictional surface layer. A simple argument based on conservation of vorticity and energy was given to explain the steady state flow pattern. This analysis shows that the velocity of the eastward jet varies as $\tau^{2/3}$ (τ is the meridional wind stress), and that its latitude varies as $\tau^{1/3}$ and so is relatively insensitive to the wind speed.

While we believe the dynamics discussed above are important to a thorough description of the countercurrents in the Atlantic and Pacific Oceans, they appear to play a more fundamental role in explaining the flow pattern observed in the equatorial Indian Ocean. Observations in the western Indian Ocean during the southwest monsoon (Taft and Knauss, 1967, Netherlands Meteorologisch Instituut, 1952) have shown weak eastward flow extending north of the equator and then giving way to westward currents that have a maximum at 3N (cf. Fig 18). During the northerly wind regime (i.e. the northwest monsoon) the reverse pattern is observed over the central Indian Ocean, with westward currents south of the equator.

Stability. The zonal wind cases showed no evidence of instability (e.g., meanders of the undercurrent) whatsoever. The current system associated with a south wind was seen to be barotropically unstable, the instability having a regular wavelike form in the zonal direction with a wavelength of 950 km and a period of 29 days. The interpretation of the standing wave pattern which arises in the southeast wind case is less straightforward. It appears that by absorbing energy from the mean flow the initial western boundary reflection maintains its amplitude at all longitudes. This contrasts with the linear response, in which this Bessel function shaped mode collapses toward the western side.

These results are in agreement with the finding of Philanders's (1976, 1978a) stability studies: the undercurrent itself is stable but the entire equatorial current-countercurrent system may be unstable—in particular, because of the large shears between the westward flow near the equator and the eastward flow in the North Equatorial Countercurrent. This instability is the probable source of the waves observed on the front between the Countercurrent and the South Equatorial Current in the Pacific (Legeckis, 1977). Those waves have a wavelength of about 1000 km and 25 day period.

The southeast case has an undercurrent which meanders in space but is steady in time. In the course of reaching this steady state, the undercurrent exhibited time variations not unlike those observed during GATE (Düing *et al.*, 1975). Rather than being an instability, it is possible that the observed meanders are part of the ocean's adjustment to changes in the wind (cf Hallock, 1977), though the data record is too short to be conclusive.

Acknowledgments. I am grateful to my advisor, Professor J. G. Charney, for his guidance and encouragement. Several valuable conversations with Professor H. M. Stommel are acknowledged. Professors Eugenia Rivas and Moshe Israeli provided much useful advice about numerical methods. My special thanks to Dr. E. S. Sarachik, for providing invaluable criticisms throughout the course of this work.

The computations were carried out at the Goddard Institute for Space Studies. My thanks to Dr. Robert Jastrow and Dr. Milton Halem for making this possible. This work was supported by NASA Grant NGR 22-009727 to the Massachusetts Institute of Technology.

REFERENCES

Cane, M. A. 1975. A study of the wind-driven ocean circulation in an equatorial basin, Ph.D. thesis, Massachusetts Institute of Technology, 372 pp.

—— 1977. The response of an equatorial ocean to simple wind stress patterns: I. Model formulation and analytic results. J. Mar. Res., *37*, 233–252.

Cane, M. A. and E. S. Sarachik. 1976. Forced baroclinic ocean motions: I. The linear equatorial unbounded case. J. Mar. Res., *34*, 629–665.

—— 1977. Forced baroclinic ocean motions: II. The linear equatorial bounded case. J. Mar. Res., *35*, 395–432.

—— 1979. Forced baroclinic ocean motions: III. The linear basin case. J. Mar. Res., *37* (this issue).

Charney, J. G. 1955. The generation of ocean currents by wind. J. Mar. Res., *14*, 477–498.

—— 1960. Non-linear theory of a wind-driven homogeneous layer near the equator. Deep Sea Res., *6*, 303–310.

Charney, J. G. and S. Spiegel. 1971. Structure of wind driven equatorial currents in homogeneous oceans. J. Phys. Oceanogr., *1*, 149–160.

Düing, W., P. Hisard, E. Katz, J. Krauss, J. Meincke, K. Moroshkin, G. Philander, A. Rybnikov, and K. Voigt. 1975. Meanders and long waves in the Equatorial Atlantic. Nature, *257*, 280–284.

Fofonoff, N. P. and R. B. Montgomery. 1955. The equatorial undercurrent in the light of the vorticity equation. Tellus, *7*, 518–521.

Hallock, Z. R. 1977. Wind forced equatorial waves in the Atlantic Ocean. Ph.D. Thesis, University of Miami.

Kendall, T. R. 1970. The Pacific Equatorial Countercurrent. International Center for Environmental Research, Laguna Beach, CA., 19 pp.

Knauss, J. A. 1966. Further measurements and observations on the Cromwell Current. *J. Mar. Res., 24,* 205–240.

Knox, R. A. 1976. On a long series of measurement of Indian Ocean equatorial currents near Addu Atoll. Deep Sea Res., *23,* 211–221.

Legeckis, R. 1977. Long waves in the eastern equatorial Pacific; a view from a geostationary satellite. Science, *197,* 1179–1181.

Lorenz, E. N. 1971. An N-cycle time-differencing scheme for stepwise numerical integration. Mon. Wea. Rev., *99,* 644–648.

Mesinger, F. 1972. A method for construction of second-order accuracy difference schemes permitting no false two-grid interval wave in the height field. Tellus, *XXV,* 444–457.

Montgomery, R. B. and Palmen, E. 1940. Contribution to the question of the Equatorial Counter Current. J. Mar. Res., *3,* 112–133.

Nederlands Meteorologisch Instituut, 1952. Indische Ocean oceanografische en meteorologische gegevens. Publ. no. 135, 2nd ed., *Vol. 1,* 31 pp., *Vol. 2,* 24 charts.

O'Brien, J. J. and H. E. Hurlburt. 1974. Equatorial jet in the Indian Ocean: Theory. Science, *184,* 1075–1077.

Orszag, S. and M. Israeli. 1974. Numerical simulation of viscous incompressible flows, *in* Annual Review of Fluid Mechanics, *Vol. 6,* Palo Alto, Annual Reviews, 371 pp.

Philander, S. G. H. 1971. The equatorial dynamics of a shallow homogeneous ocean. Geophys. Fluid Dyn., *2,* 219–245.

—— 1973. The equatorial thermocline. Deep Sea Res., *20,* 69–86.

—— 1976. Instabilities of zonal equatorial currents. J. Geophys. Res., *81,* 3725–3735.

—— 1978a. Instabilities of zonal equatorial currents: II. J. Geophys. Res., submitted.

—— 1978b. Variability of the tropical oceans Dyn. Atmos. Oceans, submitted.

Robinson, A. R. 1966. An investigation into the wind as the cause of the Equatorial Undercurrent. J. Mar. Res., *24,* 179–204.

Taft, B. A. and J .A. Knauss. 1967. The Equatorial Undercurrent of the Indian Ocean as observed by the Lusiad expedition. Bull. Scripps Inst. Oceanogr., *9.*

Taft, B. A., B. Hickey, C. Wunsch and D. Baker. 1974. The Cromwell Current at 150°W. Deep Sea Res., *21,* 403–430.

Wyrtki, K. 1973. An equatorial jet in the Indian Ocean. Science, *181,* 262–264.

Printed in U.S.A. for the Sears Foundation for Marine Research,
Yale University, New Haven, Connecticut, 06520, U.S.A.
Van Dyck Printing Company, North Haven, Connecticut, 06473, U.S.A.

Received: *17 June, 1977;* revised: *22 November, 1978.*

Paper 9

Reprint from Journal of Marine Research, Volume 37, 2, 1979.

Forced baroclinic ocean motions, III: The linear equatorial basin case

by Mark A. Cane[1] and E. S. Sarachik[2]

ABSTRACT

Previous work on the linear spin-up of an equatorial ocean is extended to include the specific effects of the north-south extent of the basin, thus allowing a detailed comparison of analytic spin-up theory with numerical calculations.

North-south modes in a β-plane equatorial basin are solved for both numerically and analytically. Simple approximations are developed and the modes are classified by the location of the turning points relative to the zonal walls. The modes are illustrated for three cases: A: a symmetric basin whose zonal walls are distant from the equator (compared to the equatorial radius of deformation); B: a symmetric basin whose zonal walls are close to the equator; and C: an asymmetric basin, one wall of which is close to the equator and one far.

Spin-up in response to x-independent winds in each of these basins is then calculated in terms of four elements: the x-independent response to the wind; the eastern boundary response consisting of Rossby and Rossby-Kelvin modes needed to bring the u field to zero; the western boundary response consisting of Kelvin waves needed to bring the u field to zero; and the western boundary layer consisting of trapped short Rossby waves needed to bring u to zero and meridionally redistribute mass. These elements are combined to describe how the steady (Sverdrup) solutions are approached. Special attention is paid to the fast planetary response: the equatorially confined Kelvin mode, the exponential anti-Kelvin mode with maximum amplitude on the zonal walls, and the Rossby-Kelvin mode, the latter two being modes not present on a meridionally open basin.

Finally, a numerical model is run to illustrate spin-up for various winds in each of the basins, A, B and C.

1. Introduction

This paper is the third of a series describing the linear response to simple wind stress patterns of an equatorial ocean described by the baroclinic shallow water equations.

In the first paper (Cane and Sarachik, 1976, henceforth called I), we described

1. Laboratory for Atmospheric Sciences, NASA Goddard Space Flight Center, Greenbelt, Maryland, 20770, U.S.A.

2. Center for Earth and Planetary Physics, Harvard University, Cambridge, Massachusetts, 02138, U.S.A.

the response of an unbounded equatorial ocean to zonal and meridional wind stresses switched on in time and space. The dispersive properties of these forced wave motions were clarified by investigating the dispersive properties of a closely analogous system: the barotropic vorticity equation forced by a switched on wind stress curl. It was found that the asymptotic westward response occurred behind fronts propagating with the long Rossby nondispersive wave speed; and that asymptotic eastward motions occurred only at forcing discontinuities or western boundaries and consisted of short Rossby waves in the form of Bessel function "Gulf Streams" rapidly thinning with time. These results extend directly to the planetary wave part of the baroclinic case, this case being completed by the addition of eastward responses behind fronts travelling with the Kelvin wave speed. A method, based on the one of Matsuno (1966), was then introduced for calculating the unbounded baroclinic response, and this unbounded response was calculated for several unbounded zonal and meridional wind stresses. It was found that, in general, zonal winds excite a local resonant response that has the u and h fields growing linearly with time and the v field constant with time. By contrast, a meridional wind excites local u and h fields constant in time with no steady v field at all. The role of the equatorial inertia-gravity waves in setting up these unbounded responses was elucidated.

In the second paper (Cane and Sarachik, 1977, henceforth called II), we considered the effects of meridional boundaries at $x = 0$ and $x = X_B$ and described the spin-up of this zonally bounded but meridionally unbounded equatorial ocean. The method consisted of three basic steps: (i) the calculation of the fully unbounded response, as given in I; (ii) the calculation of the boundary responses needed to bring the zonal velocity of the unbounded response to zero on the boundaries; and (iii) the calculation of the boundary reflections necessary to bring to zero the zonal velocities of any of the wave responses calculated in step (ii) and any of their subsequent reflections. These boundary reflections in steps (ii) and (iii) were calculated by a method due to Moore (Moore and Philander, 1977). The method depends in an essential way on the detailed properties of the Hermite functions, which are the correct functions to use only when the basin is meridionally unbounded. The motions excited by switched on wind stresses were followed in time and the precise manner in which the wave motions conspired to produce the steady (Sverdrup) solution was studied. It was found that the approach to the steady solution is significantly impeded when Kelvin waves are excited, for these Kelvin waves induce a sloshing of mass back and forth across the basin with only slowly decreasing amplitude. It was argued in this paper that despite the unrealistic unboundedness of the basin in the meridional direction, the results obtained should be valid in the vicinity of the equator.

It is the purpose of this paper to explicitly include the boundedness of the basin in the meridional direction in the description of the spin-up process and in so doing

to decide if any essential features are introduced by these boundaries at the north and south. Thus we will solve the non-dimensional shallow water equations

$$u_t - yv + h_x = F\,H(t) \tag{1a}$$

$$v_t + yu + h_y = G\,H(t) \tag{1b}$$

$$h_t + u_x + v_y = Q \quad (t) \tag{1c}$$

subject to the boundary conditions

$$u = 0 \text{ at } x = 0 \text{ and } X_E \tag{2a}$$

$$v = 0 \text{ at } y = Y_S \text{ and } Y_N \tag{2b}$$

where in general Y_S will be taken to be south of the equator and Y_N will be taken to be north.

The imposition of meridional boundedness in condition (2b) means that the meridional eigenmodes will no longer simply be related to the Hermite functions. Hickie (1979) solved for the eigenvalues and eigenfunctions of a basin bounded in one meridional direction and found that the deviation from the Hermite-like behavior increased as the boundary approached the equator. As we will see, Hickie's assumption of unboundedness in one direction eliminates some important solutions (we will call them anti-Kelvin waves) that were previously noted by Moore (1968) and Mofjeld and Rattray (1971) in their discussions of free modes of an equatorial basin.

In what follows, the basic method of calculating the time dependent response is essentially the same as the method of Π and can be summarized by noting the equivalent 3-step process: (i) the zonally unbounded, but meridionally bounded, response is calculated; (ii) the eastern and western boundary response needed to bring to zero the zonal velocity field of the zonally unbounded response of step (i) is calculated; (iii) the eastern and western boundary response needed to bring to zero the zonal velocity fields of any boundary responses emitted subsequently as part of step (ii) is calculated. It should be noted that the inclusion of northern and southern boundaries complicates the analytic problem significantly, yet resolution of this complication is essential before analytic results can be compared to, say, numerical simulations.

The plan of this paper is as follows: the second section will describe the eigenvalues and eigenfunctions of the v equation in a meridionally bounded domain. Analytic approximations are compared in detail with numerically generated results. Special emphasis is placed on modes not present in the meridionally unbounded case, in particular the westward propagating Rossby-Kelvin, anti-Kelvin and inertia-Kelvin modes, all of which have large amplitude at the zonal boundaries.

The third section reviews notation for the vector eigenfunctions, describes their completeness and orthogonality properties, and uses them to implement step (i),

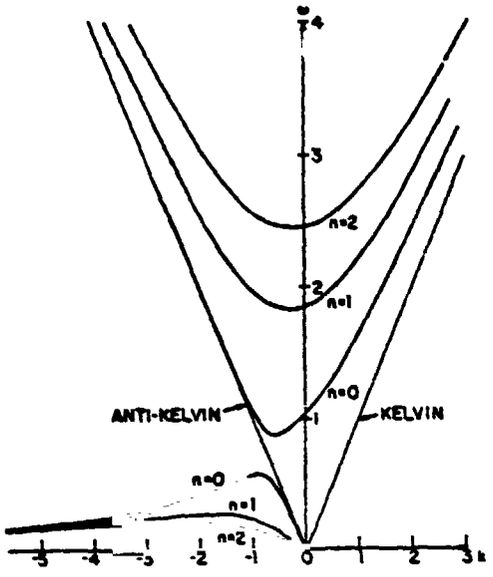


Figure 1. The dispersion relation (Eq. (5)) drawn for the eigenvalues of the [−5,1.7] case. The $n = 0$ low frequency branch is the Kelvin-Rossby mode; the $n = 0$ high frequency branch is the inertia-Kelvin mode.

the calculation of the meridionally bounded but zonally unbounded response. The fourth section shows how to calculate the eastern and western boundary response and thus how to implement steps (ii) and (iii) described above.

Section 5 will describe in some detail the time dependent response of three equatorial basins to $F = 1$ and $G = 1$ wind stresses. This will be done by presenting a linear numerical simulation of the spin-up process and comparing with the linear analytic theory developed in the earlier sections. Finally, the applicability and limits of this work to the baroclinic response of real equatorial oceans will be discussed in the last section, and our conclusions will be given.

2. The free modes of the meridional velocity equation

The homogeneous set of equations (1) with $F = G = Q = 0$ and dependence $u, v, h \sim e^{i(kx-\omega t)}$ can be reduced to a single equation for v:

$$v_{yy} + [2\mu + 1 - y^2]\, v = 0 \tag{3}$$

subject to the conditions at the northern and southern boundaries

$$v = 0 \text{ at } Y_S, Y_N . \tag{4}$$

In terms of the eigenvalues μ_n, the nondimensional dispersion relation is

$$\omega_n^2 - k^2 - \frac{k}{\omega_n} = 2\mu_n + 1 . \tag{5}$$

There are two linearly independent solutions to (3) so that the general solution of (3) may be written, in the notation of Abramowitz and Stegun (1965):

$$v(y) = U(-\mu - 1/2, \sqrt{2}\, y) + bV(-\mu - 1/2, \sqrt{2}\, y) \tag{6}$$

where $U(V)$ is the parabolic cylinder function that decays (blows up) exponentially as $y \to +\infty$.

We will take Y_y to be north of the equator and Y_s to be south so that applying the boundary condition (4) at the northern boundary yields

$$b = -\frac{U(-\mu - 1/2, \sqrt{2}\, Y_y)}{V(-\mu - 1/2, \sqrt{2}\, Y_y)}. \tag{7a}$$

Applying the boundary condition at the southern boundary yields an eigenvalue equation for the sequence of eigenvalues μ_n:

$$U(-\mu_n - 1/2, -\sqrt{2}\,|Y_s|) - \frac{U(-\mu_n - 1/2, \sqrt{2}\, Y_y)}{V(-\mu_n - 1/2, \sqrt{2}\, Y_y)} V(-\mu_n - 1/2, -\sqrt{2}\,|Y_s|) = 0. \tag{7b}$$

Since Equation (3) with boundary conditions (4) is of standard Sturm-Liouiville form, we may immediately conclude that the eigenvalues form a nonnegative increasing sequence and that the eigenfunctions are complete and orthogonal on the interval $[Y_s, Y_y]$.

Equations (6), (7a) and (7b) give the general solution to the problem, but since the content of these equations is at best opaque we will examine various approximations to the solutions with a view toward deriving analytic approximations for the eigenvalues and eigenfunctions and thus understanding them in simpler terms.

To begin with, we must recognize that the solutions U and V have Stokes lines in the complex plane so that analytic continuation across these lines changes the form of the representation. Since we will need only the representation of the solution for real positive and negative values of y, we will need only the auxiliary relations (Abramowitz and Stegun, 1965)

$$U(-\mu - 1/2, -y) = \frac{\pi V(-\mu - 1/2, y)}{\Gamma(-\mu)} + \cos \pi\mu\, U(-\mu - 1/2, y) \tag{8a}$$

and

$$V(-\mu - 1/2, -y) = (1/\pi) \sin^2 \pi\mu\, \Gamma(-\mu)\; U(-\mu - 1/2, y) - \cos \pi\mu\, V(-\mu - 1/2, y) \tag{8b}$$

both valid for $y > 0$. Using these relations in (7b), the eigenvalue equation can be written in terms of $U(y)$ and $V(y)$ with positive argument only:

$$\frac{\pi}{\Gamma(-\mu_n)\cos\pi\mu_n} + \frac{U(-\mu_n - 1/2, \sqrt{2}\,|Y_s|)}{V(-\mu_n - 1/2, \sqrt{2}\,|Y_s|)}$$
$$+ b_n\left[\frac{\sin^2\pi\mu_n}{\pi\cos\pi\mu_n}\,\Gamma(-\mu_n)\,\frac{U(-\mu_n - 1/2, \sqrt{2}\,|Y_s|)}{V(-\mu_n - 1/2, \sqrt{2}\,|Y_s|)} - 1\right] = 0$$

where b_n is given by (7a) when μ takes on one of the eigenvalues μ_n. Simplifying using the standard relation

$$\Gamma(-\mu)\,\Gamma(\mu + 1)\sin\pi\mu = -\pi \tag{9}$$

yields

$$\Gamma(\mu_n + 1)\tan\pi\mu_n - \frac{U(-\mu_n - 1/2, \sqrt{2}\,|Y_s|)}{V(-\mu_n - 1/2, \sqrt{2}\,|Y_s|)}$$
$$+ b_n\left[\frac{\cot\pi\mu_n}{\Gamma(\mu_n + 1)}\,\frac{U(-\mu_n - 1/2, \sqrt{2}\,|Y_s|)}{V(-\mu_n - 1/2, \sqrt{2}\,|Y_s|)} + 1\right] = 0\,. \tag{10}$$

In the limit that $Y_y \to \infty$, $b_n \to 0$, and using (8a), the eigenvalue equation reduces to that of Hickie (1979):

$$U(-\mu_n - 1/2, \sqrt{2}\,Y_s) = 0 \tag{11a}$$

recalling that Y_s is negative. In the opposite limit that $Y_s \to -\infty$, (10) reduces to $b_n + \Gamma(\mu_n + 1)\tan\pi\mu_n = 0$, or, again using (8a)

$$U(-\mu_n - 1/2, -\sqrt{2}\,Y_y) = 0\,. \tag{11b}$$

In the limit that both $Y_y \to +\infty$ and $Y_s \to -\infty$, (10) reduces to $\Gamma(\mu_n + 1) \cdot \tan\pi\mu_n = 0$, i.e., $\mu_n = n$ where n is a nonnegative integer. This limit represents the meridionally unbounded case treated in Papers I and II: the functions V never appear and $U(-n - 1/2, \sqrt{2}\,y)$ are simply the Hermite functions $H_n(y)\exp[-y^2/2]$ treated previously.

For the meridionally bounded case, (10) represents the general eigenvalue equation and can be solved in various analytic approximations. These approximations can be classified by the location with respect to the northern and southern boundaries of the turning points of the nth eigenfunction at $y = \pm\sqrt{2\mu_n + 1}$.

There are four cases of interest:

(i) Both turning points lie well within the basin, i.e., $2\mu_n + 1 << Y_s^2$ and Y_y^2.

Using the asymptotic forms valid for $y^2 >> 2\mu + 1$,

$$U(-\mu - 1/2, y) \to y^{\mu}\exp[-y^2/4]$$
$$V(-\mu - 1/2, y) \to \sqrt{2/\pi}\,(y)^{-\mu-1}\exp[+y^2/4] \tag{12}$$

gives

$$b_n = -\sqrt{\pi/2}\,(\sqrt{2}\,Y_N)^{2\mu_n+1}\quad e^{-Y_S^2} \tag{13}$$

and the approximate eigenvalue equation

$$\Gamma(\mu_n+1)\tan\pi\mu_n = \sqrt{\pi/2}\left[(\sqrt{2}\,|Y_S|)^{2\mu_n+1}\quad e^{-Y_S^2}\right.$$
$$\left.+\,(\sqrt{2}\,Y_N)^{2\mu_n+1}\quad e^{-Y_N^2}\right]. \tag{14}$$

Since the right-hand side of (14) is small, the eigenvalues are approximately integer and Eq. (14) can be solved approximately to yield

$$\mu_n \sim n + (2\pi)^{-1/2}(n!)^{-1}\left[\,(\sqrt{2}\,|Y_S|)^{2n+1}\,e^{-Y_S^2} + (\sqrt{2}\,Y_N)^{2n+1}\,e^{-Y_N^2}\,\right]. \tag{15}$$

The eigenfunctions are oscillatory between the turning points and decaying beyond.

(ii) The northern turning point lies well within the basin but the southern one lies well outside the basin, i.e., $Y_S^2 << 2\mu_n + 1 << Y_N^2$.

The asymptotic relations (12) are again used to give the same b_n as in (13). But for $y^2 << 2\mu_n + 1$

$$\frac{U(-\mu_n-1/2,\,y)}{V(-\mu_n-1/2,\,y)} \to \Gamma(\mu_n+1)\cot\Phi_n(y) \tag{16}$$

where, to third order in $y(2\mu_n+1)^{-1/2}$,

$$\Phi_n(y) = (\mu_n+1/2)^{1/2}\,y - (\pi/2)\,\mu_n - (1/24)\,y^3\,(\mu_n+1/2)^{-1/2}$$

The eigenvalue Eq. (10) becomes, approximately,

$$\tan\pi\mu_n = \cot\Phi_n(\sqrt{2}\,|Y_S|)$$
$$+\sqrt{\pi/2}\,\frac{(\sqrt{2}\,Y_N)^{2\mu_n+1}\quad e^{-Y_S^2}}{\Gamma(\mu_n+1)}\,[\tan\pi\mu_n\cot\Phi_n(\sqrt{2}\,|Y_S|)+1]\,. \tag{17}$$

An explicit solution to (17) is reached by first solving

$$\tan\pi\mu_n^{(0)} = \cot[(2\mu_n^{(0)}+1)^{1/2}\,|Y_S| - (\pi/2)\,\mu_n^{(0)}$$
$$-\,(1/6)\,|Y_S|^3\,(2\mu_n^{(0)}+1)^{-1/2}] \tag{18}$$

for $\mu_n^{(0)}$. (Note that $\mu_n^{(0)}$ is the eigenvalue when $Y_N \to +\infty$.)

The approximate solutions to (17) are then

$$\mu_n = \mu_n^{(0)} + \sqrt{2/\pi}\,\frac{(\sqrt{2}\,Y_N)^{2\mu_n^{(0)}+1}}{\Gamma(\mu_n^{(0)}+1)}\quad e^{-Y_S^2} \tag{19}$$

$$\cdot \left[1 + (2/\pi) \frac{Y_N}{(2\mu_n^{(0)} + 1)^{1/2}} - \frac{1}{3\pi} \frac{Y_N^3}{(2\mu_n^{(0)} + 1)^{3/2}} \right]^{-1} \cdot$$

The eigenfunctions are oscillatory from Y_S to the turning point and decaying further northward.

(iii) The southern turning point lies well within the basin but the northern one lies well outside the basin, i.e., $Y_S^2 << 2\mu_n + 1 << Y_S^2$.

Now

$$b_n = -\Gamma(\mu_n + 1) \cot \Phi_n (\sqrt{2}\, Y_N),$$

while

$$\tan \pi\mu_n = \cot \Phi(\sqrt{2}\, Y_N)$$

$$+ \sqrt{\pi/2} \frac{(\sqrt{2}\,|Y_S|)^{2\mu_n+1}\ e^{-Y_S^2}}{\Gamma(\mu_n + 1)} \left[\tan \pi\mu_n \cot \Phi_n(\sqrt{2}\, Y_N) + 1\right].$$

Again the equation

$$\tan \mu_n^{(0)} = \cot \left[(2\mu_n^{(0)} + 1)^{1/2} Y_N - (\pi/2)\, \mu_n^{(0)} - (1/6)\, Y_N^3 (2\mu_n^{(0)} + 1)^{-1/2}\right] \tag{20a}$$

is solved and the approximate eigenvalues are

$$\mu_n = \mu_n^{(0)} + \sqrt{2/\pi} \frac{(\sqrt{2}\,|Y_S|^{2\mu_n^{(0)}+1}}{\Gamma(\mu_n^{(0)} + 1)} \tag{20b}$$

$$\cdot e^{-Y_S^2} \left[1 + (2/\pi) \frac{|Y_S|}{(2\mu_n^{(0)} + 1)^{1/2}} - \frac{1}{3\pi} \frac{|Y_S|^3}{(2\mu_n^{(0)} + 1)^{3/2}} \right]^{-1}$$

The eigenfunctions are decaying south of the turning point and oscillatory from the turning point to Y_N.

(iv) Both turning points lie well outside the basin, i.e., Y_S^2 and $Y_N^2 << 2\mu_n + 1$.

Only the asymptotic relation (16) need be used for this case. We find

$$b_n = -\Gamma(\mu_n + 1) \cot \Phi_n (\sqrt{2}\, Y_N) \tag{21}$$

and the eigenvalue Eq. (10) reduces simply to

$$\sin \left[(2\mu_n + 1)^{1/2} (|Y_S| + Y_N) - (1/6)\, (2\mu_n + 1)^{-1/2} \cdot (|Y_S|^3 + Y_N^3)\right] = 0 \tag{22}$$

with solution, valid to $0(1/n^2)$,

$$\mu_{n-1} = (1/2) \left\{ \frac{n^2\pi^2}{(|Y_S| + Y_N)^2} + (1/3) \frac{|Y_S|^3 + Y_N^3}{|Y_S| + Y_N} - 1 - \frac{(|Y_S|^3 + Y_N^3)^2}{36\pi^2\pi^4} \right\}, \qquad n = 1, 2, 3, \ldots \tag{23}$$

The solution (22) and (23), to order n^2, is identical to the solution of the equations of motions without any coriolis force at all in a meridionally bounded basin; the corresponding eigenfunctions are just sines. The additional terms to 0(1) and $0(1/n^2)$ are significant corrections for the low order modes but lose their importance for the higher modes.

Finally, we note that for $2\mu_n + 1$ large (it need be no smaller than 10), there is a *uniform* approximation for the solutions of (3) (Abramowitz and Stegun, 1965)

$$U(-\mu - 1/2, y) = 2^{\mu/2}\,\Gamma\left(\frac{\mu+1}{2}\right)\left(\frac{t}{\xi^2-1}\right)^{1/4} \quad Ai(t) \tag{24a}$$

and

$$\Gamma(\mu+1)\,V(-\mu-1/2, y) = 2^{\mu/2}\,\Gamma\left(\frac{\mu+1}{2}\right)\left(\frac{t}{\xi^2-1}\right)^{1/4} \quad Bi(t) \tag{24b}$$

where

$$\xi = y(2\mu+1)^{-1/2},\ t = (4\mu+2)^{2/3}\,\tau$$

and

$$\tau = -\left(\frac{3}{2}\ \theta_1\right)^{2/3} \qquad \xi \leqslant 1\,,$$

$$\tau = \left(\frac{3}{2}\ \theta_2\right)^{2/3} \qquad \xi \geqslant 1$$

with

$$\theta_1 = (1/4)\ \arccos \xi - (1/4)\,\xi\sqrt{1-\xi^2}$$

$$\theta_2 = (1/4)\,\xi\sqrt{\xi^2-1} - (1/4)\ \operatorname{arccosh} \xi\,.$$

(7a) and (10) then become

$$b_n = -\,\Gamma(\mu_n+1)\,\frac{Ai[t_n(\sqrt{2}\,Y_N)]}{Bi[t_n(\sqrt{2}\,Y_N)]} \tag{25}$$

and

$$\tan \pi\mu_n \left[\ 1 - \frac{Ai[t_n(\sqrt{2}\,Y_N)]\,Ai[t_n(\sqrt{2}\,|Y_S|)]}{Bi[t_n(\sqrt{2}\,Y_N)]\,Bi[t_n(\sqrt{2}\,|Y_S|)]}\ \right] - \left[\frac{Ai[t_n(\sqrt{2}\,|Y_S|)]}{Bi[t_n(\sqrt{2}\,|Y_S|)]}\right] - \left[\frac{Ai[t_n(\sqrt{2}\,Y_N)]}{Bi[t_n(\sqrt{2}\,Y_N)]}\right] = 0\,. \tag{26}$$

It can be checked, by using the appropriate asymptotics of the Airy functions, that cases (i), (ii), (iii) and (iv) can be derived as the appropriate low order limits of (25) and (26).

We will illustrate these results by comparing these analytic approximations with

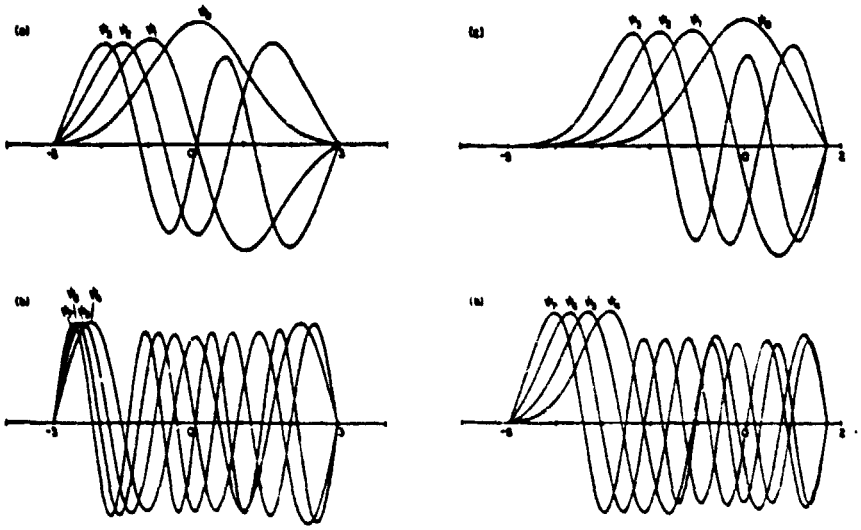


Figure 2. (left). Normalized eigenmodes of the v equation (3) for a $[-3,3]$ basin. (a) Modes $n = 0$ to 3; (b) Modes $n = 4$ to 7.

Figure 3. (right). Normalized eigenmodes of the v equation (3) for an asymmetric $[-5,1.7]$ basin. (a) Modes $n = 0$ to 3; (b) Modes $n = 4$ to 7.

results obtained by solving (3) and (4) by direct numerical means. We will consider three separate basins, one basin symmetric about the equator whose walls are relatively far from the equator, $[-3,3]$, one asymmetric basin whose northern wall is relatively close to the equator at approximately the same position as the Gulf of Guinea, $[-5,1.7]$, and a symmetric basin both of whose walls are close to the equator, $[-1.7,1.7]$.

a. Basin $[-3,3]$

The first eight eigenfunctions are shown in Figure 2. The first four have $\sqrt{2\mu + 1} < 3$ so their turning points are inside the basin. In each of these first four eigenfunctions we see evidence of decay toward the northern and southern boundaries with oscillatory behavior equatorward of this decay. Starting with the fifth eigenfunction, the turning points are outside the basin and no evidence of decay toward the walls is seen. In accordance with the properties of the Airy and Bairy functions in (24), we notice that the envelope of the oscillatory part of the eigenfunctions decays toward the equator.

We can compare the numerically obtained eigenvalue with the two approximations relevant to this case—(i) and (iv).

The approximation (15) for the lowest two eigenvalues is quite good because the exponential decay of the eigenfunction toward the walls is quite rapid. The approximation (23) is less good because, while the turning points are reasonably far outside

	Numerical	Eq. (15)	Eq. (23)
μ_0	.00039	.00042	invalid
μ_1	1.006	1.008	invalid
μ_2	2.041	2.068	invalid
μ_3	3.164	3.406	invalid
μ_4	4.454	invalid	4.263
μ_5	5.973	invalid	5.821
μ_6	7.753	invalid	7.633
μ_7	9.804	invalid	9.708
μ_8	12.129	invalid	12.053

the basin, the eigenvalue is not yet very large compared to unity. The approximations are least good when the turning point is near the wall.

b. Basin [*−5,1.7*]

The first eight eigenfunctions are shown in Figure 3. In this case, only the lowest eigenfunction has both turning points inside the basin. Eigenfunctions 1 through 7 have the northern turning point outside the basin and the southern one inside the basin. Eigenfunction 8 (not shown) has its southern turning point right at the boundary, and eigenfunctions 9 and above have both their turning points outside the basin.

	Numerical	Eq. (20a)	Eq. (20b)	Eq. (15)	Eq. (23)
μ_0	.046	invalid		.053	invalid
μ_1	1.220	1.252	1.252	1.308	invalid
μ_2	2.532	2.542	2.542	invalid	invalid
μ_3	3.943	3.947	3.947	invalid	invalid
μ_4	5.424	5.526	5.426	invalid	invalid
μ_5	6.957	6.956	6.963	invalid	invalid
μ_6	8.541	8.525	8.645	invalid	invalid
μ_7	10.209	10.124	invalid	invalid	invalid
μ_8	12.025	11.749	invalid	invalid	11.343
μ_9	14.047	invalid	invalid	invalid	13.487

We see that in all cases the eigenvalues are reasonably approximated (to no worse than a few percent) by the expression (20), (15) or (23) in the appropriate regions of validity. The one exception is the eigenvalue whose eigenfunction has a turning point at the wall and this one is approximated to 10%.

c. Basin [*−1.7,1.7*]

Only the lowest eigenfunction in this small basin case has both turning points inside the basin, all others having both turning points outside the basin. The lowest

eigenfunction looks gaussian, all higher ones are indistinguishable from sines and cosines.

	Numerical	Eq. (15)	Eq. (23)
μ_0	.103	.107	invalid
μ_1	1.603	invalid	1.655
μ_2	3.793	invalid	3.808
μ_3	6.797	invalid	6.803
μ_4	10.645	invalid	10.648
μ_5	15.344	invalid	15.346
μ_6	20.895	invalid	20.896
μ_7	27.299	invalid	27.300
μ_8	34.557	invalid	34.558

The approximations for the lowest eigenvalue given by (15) is quite good, while the approximation (23), valid when both turning points are outside the basin, is essentially exact for the higher eigenvalues.

We can summarize the solutions to the v equation by noting simply that those low order modes whose turning points both lie within the basin resemble the unbounded solutions in the sense that the eigenvalues are exponentially close to integers and the eigenfunctions are very close to the unbounded solutions (Hermite functions) except that they are brought to zero within a local Rossby radius of deformation of the boundaries. Those high order modes whose turning points both lie outside the basin are oscillatory in structure, being quantized by the north and south walls, and to lowest order (n^2) resemble modes in a non-rotating system, the effects of β entering only to order 1.

While the low order v modes resemble their unbounded counterparts, the enforced vanishing at the boundaries leads to interesting behavior of the associated zonal velocity and height fields.

Corresponding to each eigenfunction

$$v_n = U(-\mu_n - 1/2, \sqrt{2}\, y) + b_n V(-\mu_n - 1/2, \sqrt{2}\, y)$$

are u and h fields given by

$$u_n = \frac{i}{\omega_n^2 - k^2}\left[\omega_n y v_n - k v_n'\right]$$

$$h_n = \frac{i}{\omega_n^2 - k^2}\left[k y v_n - \omega v_n'\right] \tag{27}$$

where ω_n satisfies the dispersion relation (5).

Because the derivative of a series asymptotic to $v_n(y)$ need not be asymptotic to $\frac{dv_n}{dy}$, it proves convenient to re-express (27) in terms of U and V only. This can

be done using

$$\frac{d}{dy}\, U(-\mu-1/2, \sqrt{2}\, y) = -y\, U(-\mu-1/2, \sqrt{2}\, y) + \sqrt{2}\, \mu\, U(-\mu+1/2, \sqrt{2}\, y) \tag{28}$$

and

$$\frac{d}{dy}\, V(-\mu-1/2, \sqrt{2}\, y) = -y\, V(-\mu-1/2, \sqrt{2}\, y) + \sqrt{2}\, V(-\mu+1/2, \sqrt{2}\, y)$$

(Abramowitz and Stegun, 1965).

This yields

$$u_n(y) = i(\omega_n^2 - k^2)^{-1}[(\omega_n + k)y\, v_n(y) - \sqrt{2}\, k\, \mu_n \cdot U(-\mu_n+1/2, \sqrt{2}\, y) - \sqrt{2}\, k\, b_n\, V(-\mu_n+1/2, \sqrt{2}\, y)] \tag{29a}$$

$$h_n(y) = i(\omega_n^2 - k^2)^{-1}[(\omega_n + k)y\, v_n(y) - \sqrt{2}\, \omega_n\, \mu_n \cdot U(-\mu_n+1/2, \sqrt{2}\, y) - \sqrt{2}\, \omega_n\, b_n\, V(-\mu_n+1/2, \sqrt{2}\, y)] \tag{29b}$$

where ω_n and k must satisfy the dispersion relation (5).

We can illustrate the use of (29) by examining the behavior of the u and h fields corresponding to the lowest eigenvalue $\mu_0 << 1$. Consider, for convenience, a symmetric basin extending from $Y_s = -L$ to $Y_y = L$ where $L \geqslant 1$. Then

$$\mu_0 \cong 2\pi^{-1/2}\, L \exp[-L^2] \qquad \text{(using Eq. (15))}$$

and

$$b_0 \cong -(1/2)\pi\, \mu_0 \qquad \text{(using Eq. (13))}$$

and

$$\omega_0^2 - k^2 - \frac{k}{\omega_0} = 1 + 2\mu_0 \qquad \text{(see Fig. 1).}$$

Consider the region near ω, $k \sim 0$, where $\omega_0 \sim -k(2\mu_0 + 1)^{-1}$, so that we are on the "Rossby-Kelvin" part of the dispersion curve (the reason for this nomenclature will soon be apparent). The μ_0 and h_0 fields can thus be written

$$u_0 = \frac{i}{2\omega}\left[\, y\, v_0(y) - (1/\sqrt{2})U(-\mu_0+1/2, \sqrt{2}\, y) + \frac{\pi}{2\sqrt{2}}\, V(-\mu_0+1/2, \sqrt{2}\, y)\right] \tag{30a}$$

$$h_0 = \frac{i}{2\omega}\left[\, y\, v_0(y) + (1/\sqrt{2})U(-\mu_0+1/2, \sqrt{2}\, y) - \frac{\pi}{2\sqrt{2}}\, V(-\mu_0+1/2, \sqrt{2}\, y)\right]. \tag{30b}$$

As y gets large compared to one, the V term dominates and

$$\left.\begin{aligned} u_0 &\to \frac{i}{4\omega}\sqrt{\pi}\exp[y^2/2] \\ h_0 &\to -\frac{i}{4\omega}\sqrt{\pi}\exp[y^2/2] \end{aligned}\right\} \; y >> 1\,. \tag{31a}$$

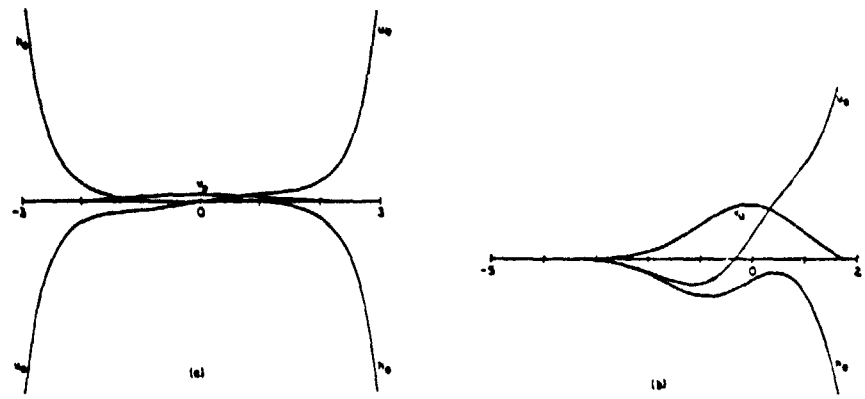


Figure 4. u, v, and h fields for the $n = 0$ Rossby-Kelvin wave. (a) In a basin $[-3,3]$, $\omega = +.5$, $k = -.501$; (b) In a basin $[-.5,1.7]$, $\omega = .5$, $k = -.603$.

For y large and negative, use of (8) with $\mu_0 \sim 0$ gives

$$\left.\begin{aligned} u_0 &\to -\frac{i}{4\omega}\sqrt{\pi}\exp\,[y^2/2] \\ h_0 &\to \frac{i}{4\omega}\sqrt{\pi}\exp\,[y^2/2] \end{aligned}\right\} \; y << -1\,. \tag{31b}$$

We see that the u and h field corresponding to the lowest eigenfunction correspond to a westward propagating anti-symmetric Kelvin wave on the northern and southern wall. Figure 4a shows the u_0 and h_0 field that belongs to the lowest eigenfunction in a $[-3,3]$ basin. The amplitude of u_0 and h_0 is $0(1/\omega)$ with respect to v_0 and so is increasingly dominant at low frequency.

Using the same sort of reasoning, we can easily show that even when the frequency and wave number are not small, those parts of the $n = 0$ dispersion curve, both Rossby-Kelvin and inertia-Kelvin, that lie near $\omega + k = 0$ have their u and h fields decay exponentially away from the walls and are thus "Kelvin-like."

As we go to higher mode numbers n, b_n in (29) stays exponentially small while μ_n increases at least as n (and ultimately as n^2) so that while neither u nor h goes to zero on the boundaries, the V term no longer dominates and the behavior is not Kelvin-like near the boundaries. (Fig. 5 shows the $n = 1$ long Rossby mode in a symmetric basin.)

The two solutions left out of our catalog are those with $v \equiv 0$ and these correspond to:

(a) An eastward propagating symmetric equatorial Kelvin wave

$$u_E^- = h_E^- \sim \exp\,[-y^2/2], \qquad \omega = k$$

with maximum amplitude on the equator.

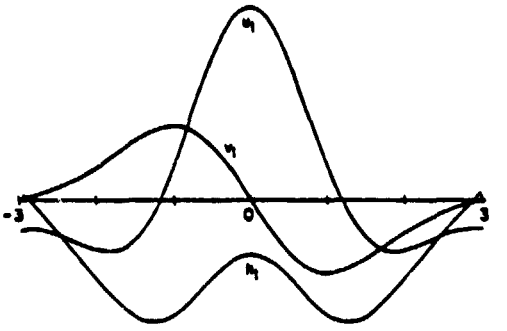


Figure 5. u, v, and h fields for the $n = 1$ Kossby mode, in a basin $[-3,3]$ $\omega = +.2$, $k = -.690$.

Because the basin is bounded, another solution exists, namely,
(b) A westward propagating symmetric "anti-"Kelvin wave

$$u_K^+ = h_K^+ \sim \exp[y^2/2], \qquad \omega = -k \tag{32}$$

with maximum amplitude at the northern and/or southern boundary.

As Moore (1968) has pointed out, the dimensional form of (32) near the northern boundary, say, can be written

$$\exp\left[\frac{1}{2}(Y_y - \eta)^2 \frac{\beta}{c}\right] = \exp\left[\frac{\beta Y_y^2}{2c}\right] \exp[-\eta/L_R] \exp\left[\frac{\eta^2\beta}{2c}\right] \tag{33}$$

where η is the distance from the northern wall and $L_R = \frac{c}{\beta Y_y}$ is the Rossby radius of deformation characteristic to the northern wall. Since $\frac{\eta^2\beta}{c} << 1$, (33) shows that the function $\exp[y^2/2]$ does indeed behave, approximately, like our usual concept of a coastal Kelvin wave, decaying exponentially away from the boundary with scale L_R.

Sums and differences of the (anti-symmetric) Rossby-Kelvin wave (Eq. (31)) and the (symmetric) anti-Kelvin wave (Eq. (32)) can be taken to produce isolated coastal Kelvin waves on either the northern or the southern boundaries in this symmetric case.

The situation is slightly different in the asymmetric case. If, say, the southern wall is much further from the equator than the northern wall, i.e., $|Y_s| >> Y_y > 1$, then the relation between b_0 and μ_0 changes to $b_0 = -\pi\mu_0$. The northern wall behavior in (30) is still dominated by the V term yielding

$$\left.\begin{aligned} u_0 &\to \frac{i}{2\omega}\sqrt{\pi}\exp[y^2/2] \\ h_0 &\to -\frac{i}{2\omega}\sqrt{\pi}\exp[y^2/2] \end{aligned}\right\} \text{ as } y \to Y_y .$$

Now, however, analysis of (30) for $y < 0$, using (8), gives a cancellation in which V no longer enters and the u_0 and h_0 fields south of the equator decay exponentially to the southern wall. (This behavior is illustrated in Figure 4b for a [−5,1.7] basin.)

Isolated northern wall-trapped Kelvin waves can thus, in this asymmetric case, be constructed out of the $n = 0$ Rossby-Kelvin mode alone. Isolated Kelvin waves trapped to the southern boundary can be constructed out of an anti-Kelvin wave (32), with the northern branch subtracted away by an equal amount of $n = 0$ Rossby-Kelvin mode.

3. Review and notation

a. Free modes

In the last section we have considered the eigenfunctions of the v equation (3) subject to zero boundary conditions on the north and south boundaries. If we define the normalized version of these eigenfunctions as

$$\psi_n(y) = v_n(y)\left[\int_{Y_s}^{Y_s} v_n^2(y)\,dy\right]^{-1/2}$$

then the normalized vector of the Fourier components of the free solutions to (1) subject to the boundary conditions (2a) (but not (2b)) can be written

$$[u,v,h] = \Phi_{n,j}(k,y)\,e^{i(kx-\omega_{n,j}(k)t)} \tag{34}$$

where n labels the mode number 0, 1 . . . . ∞ and j represents one of the 3 solutions to the dispersion relation (5).

Let us introduce the auxiliary vectors, as in I and II;

$$\mathbf{W}_n(y) = (y\,\psi_n(y),\, 0,\, -\psi_n'(y)) \tag{35a}$$

$$\mathbf{V}_n(y) = (0,\, \psi_n(y),\, 0) \tag{35b}$$

and

$$\mathbf{M}_n(y) = (-\psi_n'(y),\, 0,\, y\psi_n(y)) \tag{35c}$$

and the vector product

$$[\mathbf{A}(y),\mathbf{B}(y)] = \int_{Y_s}^{Y_s} [A_1^*\,B_1 + A_2^*\,B_2 + A_3^*\,B_3]\,dy$$

in terms of which it is easy to derive the following properties of the auxiliary functions directly from the defining Eq. (3):

$$\begin{aligned} &[\mathbf{W}_m,\mathbf{W}_n] = (2\mu_n + 1)\,\delta_{mn} && [\mathbf{W}_m,\mathbf{M}_n] = \delta_{mn} \\ &[\mathbf{M}_m,\mathbf{M}_n] = (2\mu_n + 1)\,\delta_{mn} && [\mathbf{M}_m,\mathbf{V}_n] = 0 \\ &[\mathbf{V}_m,\mathbf{V}_n] = \delta_{mn} && [\mathbf{W}_m,\mathbf{V}_n] = 0\,. \end{aligned} \tag{36}$$

The free solutions (34) then become

$$\Phi_{n,j}(k,y) = [\omega_{n,j}(k)\mathbf{W}_n + k\,\mathbf{M}_n(y) - i(\omega_{n,j}^2 - k^2)\,\mathbf{V}_n(y)]\,N_{n,j}^{-1} \tag{37}$$

for $n = 0, 1 \ldots$ and $j = 1, 2, 3$, where the normalization factor $N_{n,j}$ is given by

$$N_{n,j}^2 = (2\mu_n + 1)\,(\omega_{n,j}^2 + k^2) + 2k\,\omega_{n,j} + (\omega_{n,j}^2 - k^2)^2\,.$$

These functions are complete, as shown in the Appendix, and have the orthonormality property

$$[\Phi_{m,i}, \Phi_{n,j}] = \delta_{nm}\,\delta_{ij}\,. \tag{38}$$

In addition, it proves useful to introduce a vector which has $v = 0$ and whose u and h components are proportional to the Rossby ($\omega \sim -k(2\mu_n + 1)^{-1}$) limits of (37):

$$\mathbf{R}_n = (4\mu_n\,(\mu_n + 1))^{-1}\,[(2\mu_n + 1)\,\mathbf{M}_n - \mathbf{W}_n] \tag{39}$$

which has the following properties:

$$[\mathbf{R}_m,\mathbf{R}_n] = (2\mu_n + 1)\,(4\mu_n\,(\mu_n + 1))^{-1}\,\delta_{mn}$$

$$[\mathbf{W}_m,\mathbf{R}_n] = 0,\ [\mathbf{M}_m,\mathbf{R}_n] = 0\,. \tag{40}$$

The $v = 0$ modes discussed in Section 2 are written:

$$(u,0,h) = \Phi_{K,\pm 1}(y)\,\exp\,[ik(x \pm t)]$$

where

$$\Phi_{K,\pm 1}(y) = \mathbf{M}_K^{\pm} = 2^{-1/2}\,(\psi_K^{\pm},0,\mp\psi_K^{\pm}) \tag{41}$$

and

$$\psi_K^{\pm}(y) = a^{\pm}\,\exp\,[\pm y^2/2]$$

is normalized by $a^{\pm} = \left[\int_{Y_s}^{Y_y} \exp\,[\pm y^2/dy\right]^{-1/2}$. The $\mathbf{M}_K^{\pm}$ are normalized and mutually orthogonal. The "minus" solution is the equatorial Kelvin wave while the "plus" solution is the symmetric anti-Kelvin wave with maximum amplitude along the northern and southern boundaries.

It is easily verified that $[\mathbf{M}_K^{\pm},\mathbf{W}_n] = [\mathbf{M}_K^{\pm},\mathbf{M}_n] = 0$ so that $[\mathbf{M}_K^{\pm},\mathbf{R}_n] = 0$.

If either Y_s or Y_y recedes to infinity the anti-Kelvin wave is no longer a solution. If both Y_s and Y_y recede to infinity then the anti-Kelvin mode $\Phi_{K,+1}$, the Rossby-Kelvin mode $\Phi_{0,3}$, and the gravity Kelvin mode $\Phi_{0,1}$ are no longer solutions. The short wave part of $\Phi_{0,3}$ and $\Phi_{0,2}$ then merge to produce the "mixed" mode (cf. Matsuno, 1966).

b. The x-independent forced response

The forced response to a vector of x-independent forcings $\mathbf{F} = (F,G,Q)$ switched on at $t = 0$ is, by the methods of I:

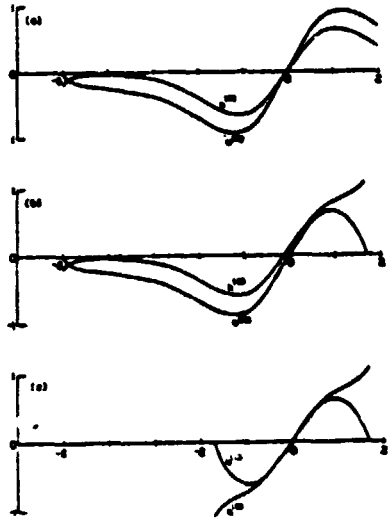


Figure 6. Zonally unbounded response to a meridional wind $G = 1$. (a) Basin $[-5,5]$ (only the part between -5 and 2 is shown—the remainder is known by symmetry); (b) Basin $[-5,1.7]$; (c) Basin $[-1.7,1.7]$.

$$\mathbf{u} = (u^{(1)}(y)t,\ v^{(1)}(y),\ h^{(1)}(y)t) + (u^{(2)}(y),\ 0,\ h^{(2)}(y)) + \mathbf{u}_I(y,t) \tag{42}$$

where

$$(u^{(1)},v^{(1)},h^{(1)}) = d_E^{-}\,\mathbf{M}_E^{-}(y) + d_E^{+}\,\mathbf{M}_E^{+} + \sum_{n=0}^{\infty} r_n\,\mathbf{R}_n - \sum_{n=0}^{\infty} (2\mu_n + 1)^{-1}\,d_n\mathbf{V}_n(y) \tag{43}$$

$$(u^{(2)},0,h^{(2)}) = \sum_{n=0}^{\infty} (2\mu_n + 1)^{-1}\,g_n\,\mathbf{W}_n(y) \tag{44}$$

and

$$\mathbf{u}_I = \sum_{n=0}^{\infty} m^{-2}\left[\mathbf{W}_n(y) + \mathbf{V}_n(y)\frac{d}{dt}\right][m^{-1}\,d_n \sin mt + g_n \cos mt] \tag{45}$$

where $m = (2\mu_n + 1)^{1/2}$.

The response $(u^{(1)}t,\ v^{(1)},\ h^{(1)}t)$ is due to zonal wind and mass forcings with

$$d_E^{\pm} = 2^{-1/2}\,(F + Q)_E^{\pm}$$

$$d_n = (yF + d\,Q/dy)_n$$

$$r_n = (dF/dy + yQ)_n - (2\mu_n + 1)^{-1}\,d_n$$

where the notation $(A)_n \equiv \int_{Y_s}^{Y_N} A\psi_n\,dy$.

The response $(u^{(2)},\ 0,\ h^{(2)})$ is d . o meridional wind forcings, where $g_n = (G)_n$.

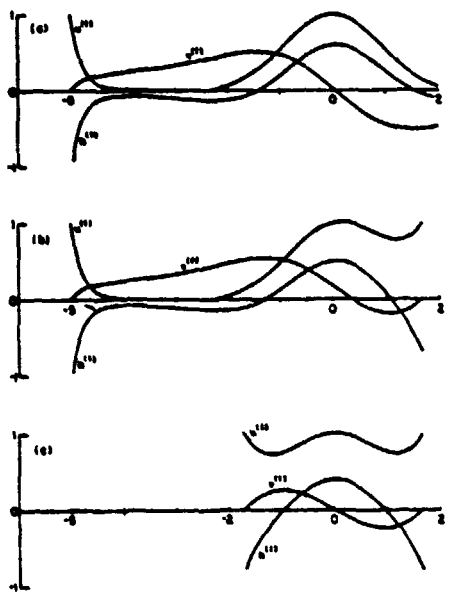


Figure 7. Same as Figure 6 but for zonal wind $F = 1$.

The response u_I is the initial intertial gravity response needed to set up the x-independent fields (see I and II for a thorough discussion).

By applying the operator $L \equiv \left(\frac{d^2}{dy^2} - y^2\right)$ with zero boundary conditions at Y_s and Y_N, it is easy to see from (44) and (3) that

$$L(-u^{(2)}/y) = L\left(\int_{Y_s}^{y} h^{(2)}\, dy\right) = G \tag{46}$$

Figure 6 shows the results of solving (46) numerically in the three basins of interest for $G = 1$.

Similarly, applying L to the $v^{(1)}$ part of (43) gives

$$L(v^{(1)}) = yF \tag{47a}$$

and in terms of this solution

$$h^{(1)} = -v_y^{(1)} \tag{47b}$$

$$u^{(1)} = F + yv^{(1)}\,. \tag{47c}$$

Figure 7 shows the results of solving (47a, b, c) in the three basins of interest for $F = 1$.

c. Examples

The modal decompositions of $u^{(1)}$ and $u^{(2)}$ are given, for five different north-south basins, in Tables 1 and 2 for the specific cases $F = 1$ and $G = 1$, respectively.

Table 1. The modal decomposition of the zonally unbounded response $u^{(\omega)}$ to $F = 1$ and its reflection at the eastern boundary. The third column lists the coefficients entering the secularly growing part of (42) and (43), and the fourth column lists the coefficients of the steady part of this response. The fifth column lists the meridionally integrated zonal mass flux for each mode entering $u^{(\omega)}$, the sum of all modes summing up to the total in the eighth column. The sixth column gives the coefficients of the eastern boundary reflection of $u^{(\omega)}$ in Eq. (54) and the seventh the zonal mass flux of each component of the reflection.

Basin	Mode	d_E^*, r_*	d_*	$U^{(\omega)}$	c^{refl}	U^{refl}	$\int_{Y_s}^{Y_n} u^{(\omega)}\,dy$
$[-\infty,\infty]$							2.122
	$K(+)$	0	—	0	0	0	
	$K(-)$	1.331	—	1.772	0	0	
	1	− .888	2.663	.295	3.550	−1.181	
	3	− .466	3.261	.032	3.727	− .253	
	5	− .331	3.646	.010	3.977	− .122	
	7	− .263	3.938	.005	4.200	− .074	
$[-5,5]$							2.504
	$K(+)$	.655	—	.429	−1.310	− .858	
	$K(-)$	1.331	—	1.773	0	0	
	1	− .888	2.663	.296	3.550	−1.182	
	3	− .466	3.259	.032	3.724	− .254	
	5	− .329	3.623	.010	3.951	− .122	
	7	− .251	3.774	.004	4.035	− .073	
$[-3,3]$							2.874
	$K(+)$	.931	—	.867	−1.862	−1.733	
	$K(-)$	1.328	—	1.764	0	0	
	1	− .831	2.502	.257	3.333	−1.033	
	3	− .315	2.311	.014	2.623	− .116	
	5	− .121	1.566	.001	1.687	− .015	
	7	− .051	1.502	$\sim 10^{-4}$	1.104	− .022	
$[-1.7,1.7]$							2.883
	$K(+)$	1.175	—	1.381	−2.350	−2.761	
	$K(-)$	1.223	—	1.496	0	0	
	1	− .326	1.372	.027	1.698	− .139	
	3	− .052	.756	$\sim 10^{-4}$	.808	− .003	
	5	− .014	.487	0	.502	0	
	7	− .006	.358	0	.364	0	
$[-5,1.7]$							2.660
	$K(+)$	.463	—	.251	− .927	− .431	
	$K(-)$	1.277	—	1.631	0	0	
	0	− .353	.384	.702	.737	−1.468	
	1	− .629	2.163	.126	2.792	− .558	
	2	− .225	1.362	.009	1.588	− .061	
	3	− .275	2.442	.009	2.718	− .086	
	4	− .166	1.964	.002	2.130	− .030	

Table 2. The modal decomposition of the zonally unbounded response $u^{(\infty)}$ to $G = 1$ and its reflection at the eastern boundary. The third column lists the coefficients $g_n = (1)_n$ in Eq. (44), the fifth column lists the meridionally integrated zonal mass flux for each mode. The fourth column lists the coefficients of the eastern boundary reflection in Eq. (55), while the sixth gives the reflected mass corresponding to it. The last column gives the constant h_n in Eq. (55).

Basin	Mode	g_n	c	$U_n^{(\infty)}$	U^{refl}	$\int_{Y_s}^{Y_N} u^{(\infty)}\,dy$	h_n
$[\infty,\infty]$						0	0
	0	1.883	0	0	0		
	2	1.331	6.390	0	0		
	4	1.153	10.249	0	0		
	6	1.053	13.602	0	0		
$[-5,5]$						0	0
	0	1.883	5.91×10^{-10}	0	0		
	2	1.331	6.390	0	0		
	4	1.153	10.249	0	0		
	6	1.041	13.457	0	0		
$[-3,3]$						0	0
	0	1.871	.0975	0	0		
	2	1.186	6.610	0	0		
	4	.728	9.353	0	0		
	6	.440	10.905	0	0		
$[-1.7,1.7]$						0	0
	0	1.623	1.750	0	0		
	2	.635	9.373	0	0		
	4	.353	14.770	0	0		
	6	.244	20.212	0	0		
$[-5,1.7]$						−1.596	.0986
	$K(+)$	—	2.167	—	1.004		
	0	1.762	.241	−.620	−.480		
	1	.348	1.822	−.217	.364		
	2	1.053	4.801	−.237	.183		
	3	.466	5.352	−.128	.168		
	4	.836	7.898	−.139	.111		
	5	.468	8.382	−.082	.099		
	6	.684	9.696	−.080	.063		
	7	.352	8.660	−.037	.043		

Table 1 (continued).

5	− .175	2.611	.002	2.786	− .033
6	− .117	2.109	.001	2.225	− .014
7	− .106	2.263	.001	2.369	− .012
8	− .060	1.511	$\sim10^{-4}$	1.562	− .003

For the $G = 1$ case, $g_n = (1)_n$, so that in basins symmetric with respect to the equator, only even (symmetric) modes enter, while in the [−5,1.7] basin, all modes enter. Since they are orthogonal to the forcing (0,G,0), note that neither the Kelvin nor the anti-Kelvin waves can enter the sum for $\mathbf{u}^{(2)}$. The meridionally integrated zonal mass flux is defined for each mode $\mathbf{u}_n^{(2)}$ of (44) as

$$U_n^{(2)} \equiv \int_{Y_s}^{Y_N} u_n^{(2)}\, dy = (1)_n (y)_n (2\mu_n + 1)^{-1}$$

where (44) and the definitions (35) have been used. Only in the asymmetric basin [−5,1.7] does $U^{(2)}$ have any meridionally integrated zonal mass flux—in the symmetric basins $(1)_n(y)_n = 0$ for each mode by symmetry considerations; Figure 6 makes this clear.

For the $F = 1$ case, the coefficients of the part of (43) that varies as t are $d_K^- = 2^{-1/2}(1)_K^-$, $d_K^+ = 2^{-1/2}(1)_K^+$, and $r_n = -(2\mu_n + 1)^{-1}(y)_n$ and are listed in the third column of Table 1. The meridionally integrated zonal mass flux corresponding to these modes are $U_K^- = 2^{-1}[(1)_K^-]^2$, $U_K^+ = 2^{-1}[(1)_K^+]^2$, and $U_n = [(y)_n]^2 [4\mu_n(\mu_n + 1)(2\mu_n + 1)]^{-1}$ and these fluxes are shown in the fifth column.[3]

It is of special interest to note that for those basins symmetric about the equator, the anti-Kelvin part of $\mathbf{u}^{(1)}$ carries more mass flux as the basin gets smaller. $\mathbf{u}^{(1)}$ in the infinite basin has no anti-Kelvin part (by the requirement of boundedness). As we go to the [−5,5] basin, the lowest few modes are essentially unchanged from the unbounded case because their turning points are still well within the basin. Essentially the only difference between the unbounded case and the [−5,5] case is the anti-Kelvin contribution, as can be seen graphically in Figure 7a. As the basin gets smaller, the anti-Kelvin mode increases in importance until we get to the [−1.7,1.7] basin where we see that essentially all the zonal mass flux in $\mathbf{u}^{(1)}$ can be accounted for solely by the Kelvin and anti-Kelvin modes.

4. The responses due to meridional boundaries

Our method, as in Papers I and II, consists of calculating the zonally unbounded response and then finding the response at the meridional boundaries $x = 0$ and X_E needed to bring the zonal velocity to zero. These boundary responses eventually reach the other boundary where they generate additional boundary responses, again those needed to bring the zonal velocity to zero. The detailed sequences of events leading to spin-up will be described in the next section. This section will show how to calculate all the needed boundary responses for use in the following section.

We should, at outset, make clear the relation between the method used in the

3. Note that the total mass flux in the zonally unbounded solution listed in the eighth column of Table 1 and the seventh of Table 2 is reached by numerically integrating $u^{(1)}$ and $u^{(2)}$, respectively. The failure of the modal sums to add to the total in the basin cases is a measure of the total incurred error and nowhere exceeds half a percent.

meridionally unbounded case in II (Moore's algorithm; Moore and Philander, 1977) and the method used here. Moore's algorithm depends on the detailed properties of the Hermite functions whereby the u_n and h_n field corresponding to meridional velocity v_n have components only involving Hermite functions $n+1$ and $n-1$. Thus at each boundary u_n is cancelled only by contributions from u_{n+1} and u_{n-1} and Moore's algorithm results. In the meridionally bounded case, however, the eigenvalues do not differ by integers so that while u_n can be expressed as parabolic cylinder functions of order $\mu_n \pm 1$, these same cylinder functions are not also parts of $u_{n\pm1}$: Moore's algorithm fails. The method we used instead is one of projections. In the meridionally unbounded case, we have already demonstrated a unique correspondence between projection coefficients and the coefficients of Moore's algorithm (Eq. (17) of II). In the meridionally bounded case only the projection method survives.

a. Western boundary response

The zonally unbounded solution (Eq. (42)) has planetary wave parts varying only as t^s ($s = 0$ or 1). The western boundary response must consist of a Kelvin wave plus a sum of terms composed of short wavelength Rossby waves: these are the only planetary modes with group velocity to the east. This response may be written, as in II:

$$\mathbf{u}^W(x,y,t) = b_K H(t-x)(t-x)^s \mathbf{M}_K{}^-(y) + \mathbf{u}^B(x,y,t) \tag{48}$$

where, to the lowest order in t/x, $\mathbf{u}^B$ is nondivergent with its *meridional* velocity component in geostrophic balance with the height field:

$$\mathbf{u}^B \equiv (u^B,v^B,h^B) = \left[-\frac{\partial}{\partial y}, \frac{\partial}{\partial x}, y \right] \left\{ \left(\frac{t}{x}\right)^{s/2} J_s(2\sqrt{xt})\,\chi(y) \right\} \tag{49}$$

with

$$\chi(y) = \Sigma b_n \psi_n(y)\,.$$

Because $\chi = 0$ at Y_S and Y_N it is clear that $\int_{Y_s}^{Y_N} u^B(x,y,t)\,dy = 0$. Thus, since $u^W + u(y)t^s = 0$ at $x = 0$, we can integrate to yield

$$b_K = -2^{1/2}[(1)_K{}^-]^{-1} \int_{Y_s}^{Y_N} u(y)\,dy \tag{50}$$

where

$$(1)_K{}^- \equiv \int_{Y_s}^{Y_N} \psi_K{}^-\,dy$$

is simply the projection of 1 on the Kelvin mode, and is known for each basin.

Thus, as in II, the Kelvin amplitude is determined by noting that all the incident zonal mass flux onto the western boundary is returned by the Kelvin wave. An ob-

vious corollary that will prove useful in what follows is that no Kelvin wave is excited off the western boundary when the incident integrated zonal mass flux vanishes.

With b_K known,

$$\chi(y) = \int_{Y_s}^{y} [u(y) + 2^{-1/2} b_K \psi_K^-(y)]\, dy \tag{51}$$

and the b_n's can be obtained by projection from the now known function $\chi(y)$.

Any Rossby, Rossby-Kelvin, or anti-Kelvin modes emitted as part of the eastern boundary response (described below) produce a western boundary response that can be calculated exactly as described above. In particular, all the meridionally integrated zonal mass flux is returned by the Kelvin wave.

b. Eastern boundary response

We have to calculate the eastern boundary response needed to bring the unbounded zonal velocity to zero and to reflect any Kelvin waves emitted by the eastern boundary.

As in II, we can use (1b) and (2a) to conclude that, asymptotically, the effect of the incident Kelvin mode and its reflections in the anti-Kelvin and Rossby modes is simply to raise the height uniformly (in y) at the eastern boundary by an amount A:

$$\mathbf{M}_K^- + a_K^+ \mathbf{M}_K^+ + \sum_{n=0}^{\infty} a_n \mathbf{R}_n = (0, 0, A)\,. \tag{52}$$

Taking projections successively with $\mathbf{M}_K^-$, $\mathbf{M}_K^+$ and $\mathbf{R}_m$ yields

$$A = 2^{1/2}/(1)_K^- \tag{53a}$$

$$a_K^+ = -A\, 2^{-1/2}\, (1)_K^+ = -\,(1)_K^+/(1)_K^-\,, \tag{53b}$$

and

$$a_n = A(y)_n = 2^{1/2}\,(y)_n/(1)_K^-\,. \tag{53c}$$

We can also see from (52) and (53) that the meridionally integrated mass flux associated with the reflected anti-Kelvin wave is in the opposite direction to the meridionally integrated mass flux associated with the incident Kelvin-wave. The Rossby waves (including the $n = 0$ Rossby-Kelvin wave) must carry off the difference of the mass fluxes due to the Kelvin and anti-Kelvin waves.

The eastern boundary response needed to bring the zonally unbounded response to a zonal current in (43) and (44), of form $u(y)t^s$ to zero will have the general form

$$\mathbf{u}^s = c_K^+ \zeta_K^s\, H(\zeta_K)\, \mathbf{M}_K^+ + \sum_{n=0}^{\infty} c_n\, H(\zeta_n)\, \{\zeta_n^s\, \mathbf{R}_n + s\, \mathbf{V}_n\} \tag{54}$$

where

$$\zeta_n = \iota + (2\mu_n + 1)\,(x - X_E)$$

and

$$\zeta_E = t + x - X_E .$$

The response to the Rossby and anti-Kelvin mode in (43), for example, is the corresponding free mode with the same structure. Thus $t\, r_n\, \mathbf{R}_n$ generates an eastern boundary response $-r_n\, \zeta_n\, H(\zeta_n)\mathbf{R}_n$ and $t d_E{}^+\, \mathbf{M}_E{}^+$ generates an eastern boundary response $-d_E{}^+\, \zeta_E\, H(\zeta_E)\mathbf{M}_E{}^+$. The eastern boundary response to the Kelvin part of (42) is found in an obvious way using Eq. (53) to find the reflected coefficients.

The eastern boundary response to the unbounded response generated by a meridional wind, (44), is a bit different. The bondary condition at $x = X_E$ is that $u^{(2)} + u^E = 0$. Since $v^{(2)} = 0$ and $s = 0$ in (54) we have that $v^{(2)} + v^E = 0$ and it then follows from (2b) that $h_y = G$. A concise statement of the preceding argument is that, at $x = X_E$,

$$(u^{(2)},0,h^{(2)}) + c_E{}^+\, \mathbf{M}_E{}^+ + \sum_{n=0}^{\infty} c_n\, \mathbf{R}_n = (0,0, \int_0^y G\, dy + h_0) \qquad (55)$$

where h_0 is as yet an unknown constant. This constant can be determined by noting that the Kelvin mode $\mathbf{M}_E{}^-$ is orthogonal to each term on the left-hand side of (55). Projecting $\mathbf{M}_E{}^-$ onto (55) and using the notation $I(y;G) = \int_0^y G(y')dy'$ gives

$$h_0 = -(I)_E{}^-/(1)_E{}^- . \qquad (56a)$$

Once h_0 is known, the c's can be found by projecting $\mathbf{M}_E{}^+$ and $\mathbf{R}_n$ onto (55):

$$c_E{}^+ = -2^{-1/2}\, (I)_E{}^+ - 2^{-1/2}\, [(1)_E{}^+]\, h_0 \qquad (56b)$$

and

$$c_n = (yI)_n - g_n\, (2\mu_n + 1)^{-1} + h_0\, (y)_n . \qquad (56c)$$

It should be emphasized that the results (48), (52) and (55) are asymptotic results and cannot be expected to hold at all times. Thus, for example, if an anti-Kelvin wave should hit the western boundary and reflect as boundary trapped modes $\mathbf{u}^B$ plus a Kelvin wave according to (48), it will take time for the Kelvin wave, whose amplitude is localized about the equator, to be produced by the anti-Kelvin wave, whose amplitude is localized near the (possibly distant) northern and southern boundaries. The necessary communication cannot be accomplished at a speed faster than the fastest wave in the problem, namely, unity. It in fact does seem to take place *at* speed unity by northward and southward propagating waves; these may be thought of as wall-Kelvin waves. Similarly, when a Kelvin wave hits the eastern boundary, the height field is raised first at the equator and then, with speed unity, to the north and south of the equator until the height is uniformly raised at the eastern boundary. The asymptotics of this process is analyzed in some detail by Anderson and Rowlands (1976).

Table 3. The eastern boundary reflection of M_E^-. The fourth column lists the coefficients appearing in Eq. (52), the fifth lists the amount of reflected meridionally integrated zonal mass flux for each mode, and the sixth lists the quantity A in Eq. (52).

Basin	Mode	Eigenvalue	a	U/U_E^-	A
$[-\infty,\infty]$					.751
	1	1	2.000	−.500	
	3	3	2.450	−.125	
	5	5	2.739	−.063	
	7	7	2.958	−.039	
[−5,5]					.751
	$K(+)$		−.492	−.242	
	1	$1+3.92\times10^{-6}$	2.000	−.500	
	3	$3+1.63\times10^{-4}$	2.448	−.125	
	5	$5+2.04\times10^{-4}$	2.721	−.062	
	7	$7+1.22\times10^{-6}$	2.843	−.036	
[−3,3]					.753
	$K(+)$		−.701	−.491	
	1	1.006	1.884	−.440	
	3	3.164	1.741	−.058	
	5	5.973	1.179	−.008	
	7	9.804	.793	−.001	
[−1.7,1.7]					.818
	$K(+)$		−.961	−.923	
	1	1.603	1.122	−.075	
	3	6.797	.618	−.002	
	5	15.344	.398	-2×10^{-4}	
	7	27.299	.293	-3×10^{-6}	
[−5,1.7]					.783
	$K(+)$		−.363	−.132	
	0	.046	.301	−.470	
	1	1.220	1.694	−.265	
	2	2.532	1.067	−.032	
	3	3.943	1.913	−.047	
	4	5.424	1.538	−.017	
	5	6.957	2.045	−.019	
	6	8.541	1.651	−.008	
	7	10.209	1.772	−.007	
	8	12.025	1.184	−.002	

c. Examples

In this subsection we will examine the eastern boundary response to an incident Kelvin wave in five basins of interest, and the boundary response to the unbounded solutions forced by $G = 1$ and $F = 1$ in these same five basins.

Table 3 lists the low order terms in Eq. (52) for the response of a Kelvin mode of

unit amplitude, M_K^-, striking the eastern boundary of five basins of different north-south extent. In the unbounded basin, the Kelvin wave reflects only in the odd n modes (whose zonal velocity is symmetric) with half of the incident meridionally integrated zonal mass flux, $U_K^- = 2^{-1/2}\,(1)_K^-$, being reflected in the $n = 1$ mode, and slowly decreasing amounts in the higher modes. Since in the unbounded case

$$\frac{U_n}{U_K^-} = -\frac{a_n^2}{4n(n+1)} = -\frac{2^{1/2}}{(1)_K^-}\,\frac{(y)_n^2}{4n(n+1)},$$

we can use the summation formula given in the Appendix of II to verify that $\sum_1^\infty U_n = -U_K^-$. We see from Table 3 that 27% of the zonal mass flux is returned in modes $n = 9$ and higher. Again, using the summation formula in the Appendix of II, we verify that

$$\sum_9^\infty \frac{U_n}{U_K^-} = -\frac{a_N^2}{4(N+1)} = -\frac{N!}{2^{N-1}\,(N+1)}\left[\left(\frac{N-1}{2}\right)!\,\right]^{-2} \qquad (57)$$

so that

$$\sum_9^\infty U_n = -.27\,U_K^-.$$

When we move to the $[-5,5]$ basin, we see that 24% of the incident Kelvin wave zonal mass flux is now returned by the anti-Kelvin wave while the amount returned by the first few Rossby waves hardly changes. In fact, we can easily verify that the anti-Kelvin mode in a basin $(-L,L)$ for $L >> 1$ reflects an amount of mass flux

$$U_K^+ = -[U_K^-]\,\sqrt{2}\,L^{-1}$$

which in turn is equal to the sum of the mass fluxes that would have been returned by all the unbounded modes whose turning point lies beyond L[4]. Thus in a bounded basin $[-L,L]$, only the modes whose turning points lie within the basin, $2\mu_n + 1 < L^2$, reflect any mass flux, while the purely oscillatory modes for which $2\mu_n + 1 > L^2$ reflect none. Thus, for the narrow basin case, $[-1.7,1.7]$, almost all the incident Kelvin mass flux is reflected by the anti-Kelvin mode.

The western boundary response to the unbounded solution $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$, in response to $F = 1$ and $G = 1$, respectively, contains a Kelvin wave whose amplitude is such as to reflect all the mass flux, according to Eq. (50). The total zonal mass fluxes are listed in Tables 1 and 2; the b_K's are then gotten from (50).

The eastern boundary response to the unbounded solution $\mathbf{u}^{(1)}$ and $F = 1$ is given in Table 1, and to the unbounded solution $\mathbf{u}^{(2)}$ for $G = 1$ in Table 2.

4. The first part of this result is obtained by asymptotically expanding the integrals in U_K^+/U_K^- in L while the second part is obtained by using (57) with $2N + 1 = L^2$ and using Stirling's formula to expand for large N.

5. Numerical simulations of time dependent response

In order to illustrate and make concrete the analytical results presented thus far, we will in this section present numerical results for the linear time-dependent response of three equatorial basins to the simple wind stress patterns $F = 1$ and $G = 1$. The first basin has its northern and southern boundaries relatively far from the equator at ± 5; the second is relatively narrow and extends from -1.7 to $+1.7$; the third is asymmetrical with its northern boundary at $+1.7$ and its southern boundary at -5. All three basins will be taken to be 10 units long in the zonal direction with the western boundary at $x = 0$ and the eastern boundary at $x = 10$. Since neither of the simple wind stresses we have chosen has curl, the steady state to which each basin will tend is simply the one in which the height field gradient balances the imposed wind stress. Thus in the $F = 1$ case, the height field in all three basins will be tilted from -5 to 5 with longitude; in the $G = 1$ case the height field will tilt uniformly with latitude from -5 to 5, -1.7 to 1.7, and -3.35 to 3.35 in the wide, narrow and asymmetric basins, respectively.

The numerical model used to simulate these ocean responses has been described elsewhere (Cane 1975, 1979). Because we ran three basins for each wind stress and each unit of nondimensional time required 60 computer time steps, practical considerations limited the total number of computer time steps tc 3840 (64 nondimensional time units) for each of the six cases.

a. The F = 1 Case

We will begin discussion of the $F = 1$ case by reviewing the meridionally unbounded situation as discussed in II. As soon as the wind begins to blow, inertia-gravity waves are excited in such a way that by one or two units of non-dimensional time, the unbounded solution $(u^{(1)}t, v^{(1)}, h^{(1)}t)$ has been fully developed. Also at $t = 0$ a full set of Rossby modes, $n = 1, 3, 5, \ldots$, begin propagating into the basin from the eastern boundary, and a Kelvin mode carrying the meridionally integrated zonal mass flux $\int_{-\infty}^{\infty} u^{(1)}t\, dy$ propagates into the basin from the western boundary. The initial solution is therefore:

$$\mathbf{u} = t\left[d_K^- \mathbf{M}_K^- + \sum_{n=1}^{\infty} r_n \mathbf{R}_n \right] + b_K(t-x)H(t-x)\mathbf{M}_K^- - \sum_{n=1}^{\infty} (2n-1)^{-1} d_n \mathbf{V}_n$$

$$+ \sum_{n=1}^{\infty} \Big\{ (-r_n + d_K^- a_n)\,(t+(2n+1)\,(x-10))\,\mathbf{R}_n + (2n+1)^{-1} d_n \mathbf{V}_n$$

$$\cdot \quad H(t+(2n+1)\,(x-10)) \Big\} \tag{58}$$

where the coefficient of the propagating Rossby modes is partly the direct eastern

boundary response to the Rossby part of $\mathbf{u}^{(1)}$ and partly the eastern boundary response to the Kelvin part of $\mathbf{u}^{(1)}$. If we note that $b_E = -\pi^{-1/4}(\pi^{1/2} + U_R)$, where $\pi^{1/2}$ and U_R are the total meridionally integrated zonal mass fluxes in the Kelvin and Rossby parts of $\mathbf{u}^{(1)}$ ($U_R = +.350$ in the unbounded case according to Table 1), and if we assume that enough time has gone by for *all* the Rossby modes to reach the point x, then (58) becomes, using (52):

$$\mathbf{u} = (0,0,x) + U_R\,\mathbf{M}_E^{-}(t-x) + 10\,\Sigma\, r_n(2n+1)\mathbf{R}_n$$

$$+ d_E^{-} \sum_{n=1}^{\infty} a_n(t+(2n+1)\,(x-10))\mathbf{R}_n\,. \tag{59}$$

The $(0,0,x)$ part of (59) has the correct slope but not the correct level to be the final steady state. The additional terms are those due to the initial Kelvin mode. The second term in particular does not stay around longer than $t = 10$—it then hits the eastern boundary producing a new series of Rossby modes. The Rossby modes, when they hit the western boundary, produce new Kelvin modes. We see therefore that while the initial solution contains enough Rossby modes to bring the height field to its correct tilt, it also initiates motions which continue to slosh mass back and forth across the basin (see II for a more complete discussion).

The initial series of reflections for the unbounded case can be described as follows. At $t = 0$ a Kelvin mode of negative amplitude carrying negative mass flux leaves the western boundary while a Rossby mode leaves the eastern boundary. The effect of these modes can be easily seen on the height field section across the equator in Figure 8. (Since the meridionally unbounded case cannot be simulated numerically, these figures were generated by summing modes.) The initial Kelvin mode, carrying negative mass flux, lowers the height field while the initial Rossby wave, also carrying negative mass flux, raises it. The flat part of the height field at $t = 4$ is simply the secularly growing part $h^{(1)}t$—it grows as if zonally unbounded because the effects of the boundaries have not yet reached it. At $t = 7.5$ the initial Kelvin and first Rossby modes meet and no secularly growing flat part of the height field remains. At $t = 10$ the initial Kelvin mode hits the eastern boundary and reflects as a Rossby mode of positive mass flux: this mode lowers the height field to its east as it propagates westward with speed 1/3. Thus the height field at the eastern boundary decreases uniformly with time until $t = 40$ when the Kelvin mode, due to the reflection of the first Rossby mode, hits and starts increasing the height field uniformly with time. At the western boundary the height field decreases uniformly during times 0 to 30 whereupon the second Rossby mode arrives producing a Kelvin mode of negative mass flux; at $t = 40$ the height field again begins to fall at the west.

The situation in the meridionally bounded basin is very different. Figure 9 shows the reflection diagram for the $[-5,5]$ basin, keeping track of only the Kelvin, anti-Kelvin and $n = 1$ Rossby modes. The initial unbounded solution $(u^{(1)}t,\ v^{(1)},\ h^{(1)}t)$

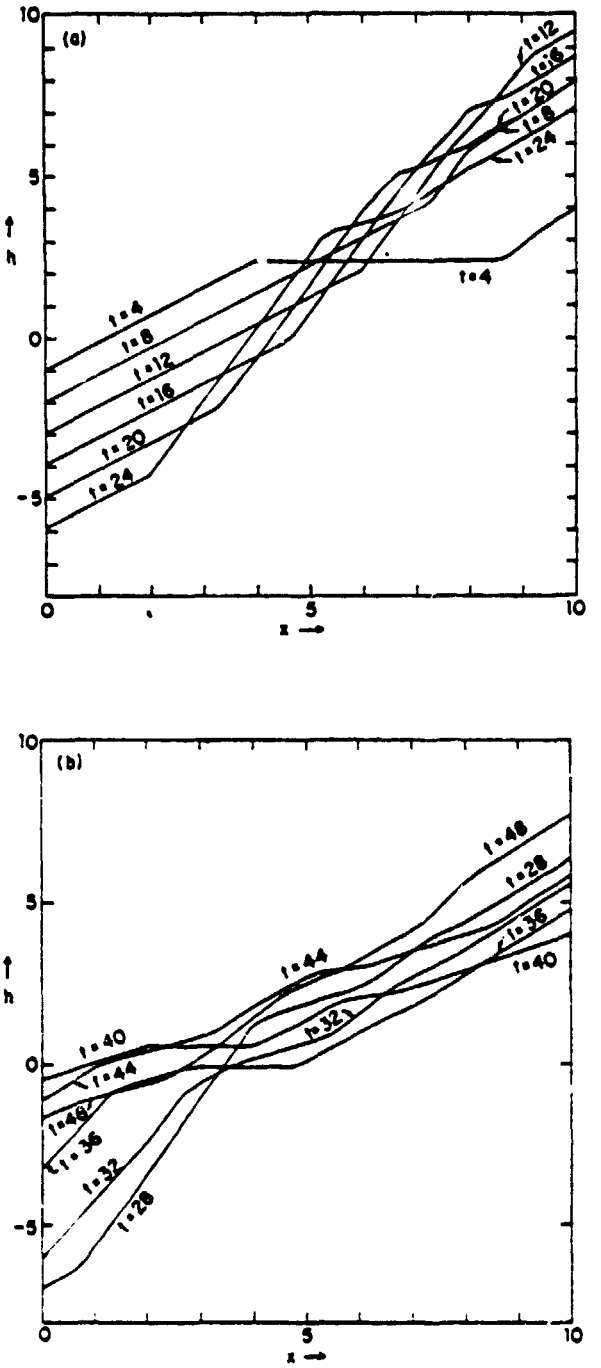


Figure 8. An equatorial section for $x = 0$ to 10 of the height field in the meridionally unbounded $F = 1$ case. (a) Times $t = 4$ to 24; (b) Times $t = 28$ to 48.

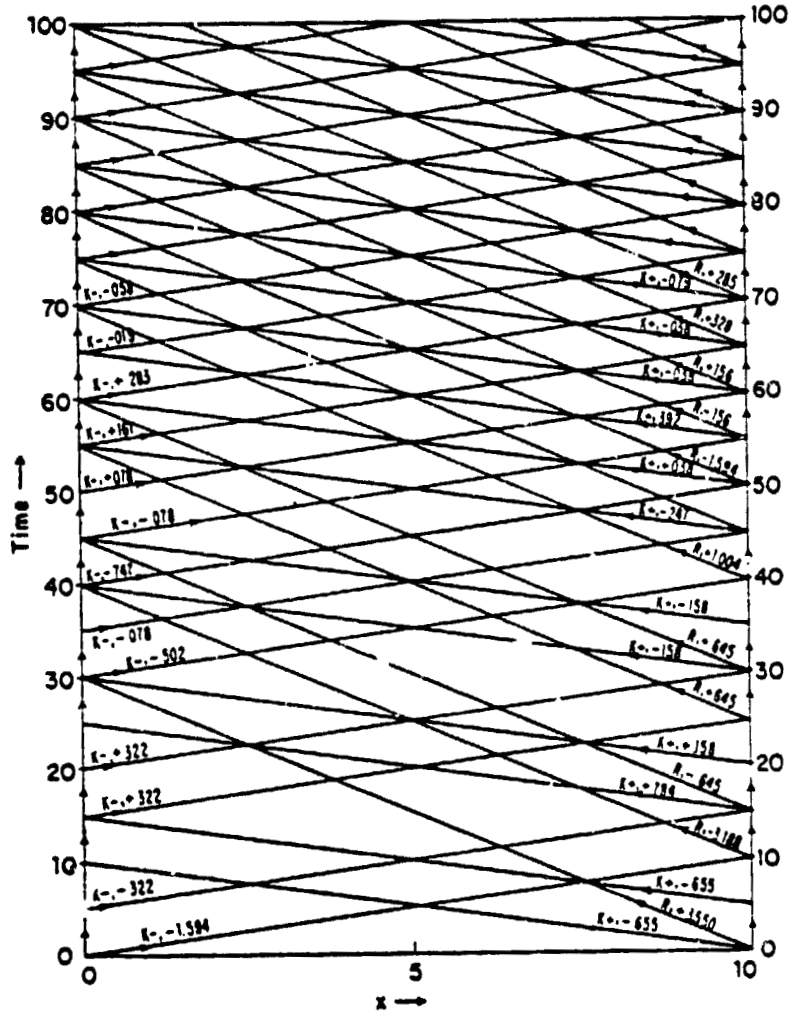


Figure 9. Reflection diagram for [−5,5] basin bounded zonally at $x = 0$ and 10 subject to uniform wind stress $F = 1$. Kelvin moder ($K-$) leave the western boundary; only the $n = 1$ Rossby (R) is kept track of leaving the eastern boundary and the anti-Kelvin mode ($K+$) is also emitted at the east. The amplitudes are written next to the mode designation.

contains an anti-Kelvin mode of amplitude .655. (Note from Table 1 that the coefficients of the Kelvin and the first four Rossby modes are essentially the same as they are in the meridionally unbounded case—the major difference between the cases is the presence of the anti-Kelvin mode.) At the western boundary a Kelvin mode reflecting the mass flux of the Kelvin and Rossby parts of $\mathbf{u}^{(1)}$ is emitted immediately while the Kelvin mode, which is to reflect the mass flux of the anti-Kelvin part of $\mathbf{u}^{(1)}$, cannot be emitted until the initial anti-Kelvin mode has communicated with the equator. As discussed in Section 4, this communication can be thought of as

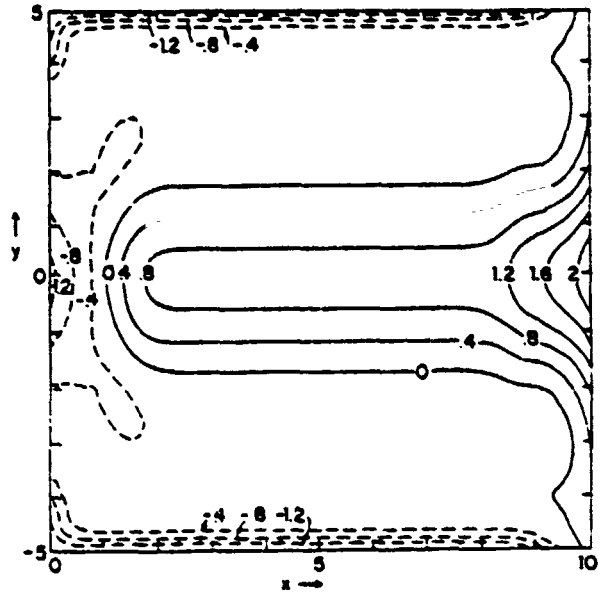


Figure 10. Contour plot of height field at $t = 2$ for $F = 1$ forcing in a $[-5,5]$ basin bounded zonally at $x = 0$ and $x = 10$.

taking place by means of equatorward travelling wall-Kelvin waves which do not reach the vicinity of the equator until $t \approx 5$—this is indicated in Figure 9 by an arrow on the western boundary leading to the emission of a Kelvin mode at $t = 5$. Similarly, at the eastern boundary, the Rossby and anti-Kelvin response to the Rossby and anti-Kelvin parts of $\mathbf{u}^{(1)}$ are emitted almost immediately while the anti-Kelvin response to the Kelvin part of $\mathbf{u}^{(1)}$ takes approximately 5 time units to develop while wall-Kelvin waves travel toward the meridional boundaries to make the connection. At later times, every anti-Kelvin mode hitting the western boundary produces a wall-Kelvin wave that travels to the equator before producing the reflected Kelvin mode and every Kelvin mode hitting the eastern boundary produces a wall-Kelvin wave that travels to the northern and southern boundaries before producing the reflected anti-Kelvin mode. The amplitudes for the various modes and reflections are given in Figure 9 until $t = 70$.

Figure 10 illustrates these initial features: it shows the height contours throughout the basin at $t = 2$. The unbounded $\mathbf{u}^{(1)}$ solution (compare to Fig. 7a) is clearly visible in the center of the basin where neither the Kelvin nor the first Rossby have yet arrived. On the western boundary, the deepening effect of the first Kelvin wave extending to $x = 2$ is seen, and near the northern and southern boundaries the anti-Kelvin part of $\mathbf{u}^{(1)}$ turning the corner as a wall-Kelvin is seen. At the eastern boundary the Rossby mode should extend 2/3 of a unit into the basin but inevitably

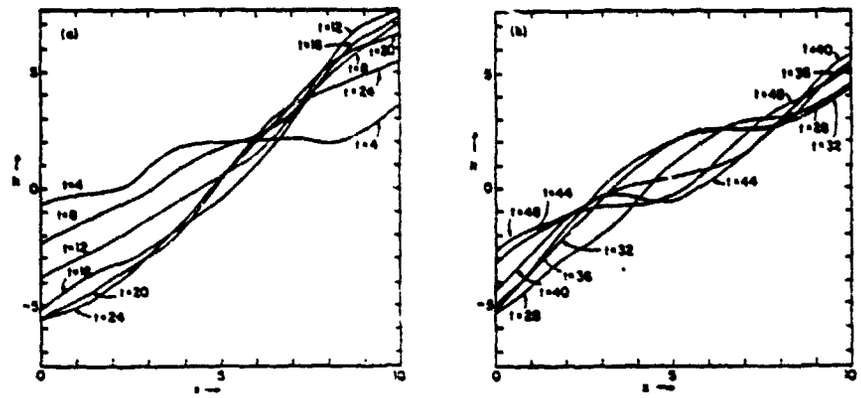


Figure 11. An equatorial section from $x = 0$ to 10 of the height field in the $[-5,5]$ $F = 1$ case. (a) Times $t = 4$ to 24; (b) Times $t = 28$ to 48.

precursors travelling as fast as speed unity produce wiggles that extend a full 2 units into the basin (see Section 3 of Paper I for a discussion of precursors and the dispersive modification of propagating Rossby fronts). The initial anti-Kelvin mode of negative amplitude (therefore positive height field) is seen propagating westward and cancelling the negative height (growing with t) anti-Kelvin mode of the unbounded solution. Finally, the wall-Kelvin (of positive height) is seen propagating toward the northern and southern boundaries. A similar contour plot 2 time units later would show these trends continuing—the unbounded solution $\mathbf{u}^{(1)}$ continuing to grow as t, the initial Kelvin mode continuing to lower the height field behind it as it propagates eastward, and the initial Rossby mode continuing to raise the height field behind it as it propagates westward, while the wall-Kelvin waves continue their journey along the eastern and western boundaries.

We can follow the progress of the system in time by examining Figure 11a, an equatorial slice of the height field from the eastern to western boundary. The deepening at the western boundary behind the initial Kelvin front proceeds faster than the comparable meridionally unbounded case in Figure 8a between times 4 and 16 because of the additional emission of the Kelvin mode of negative amplitude at $t = 5$ due to the reflection of the anti-Kelvin part of $\mathbf{u}^{(1)}$ in the basin case.[5] This more rapid deepening continues until $t = 15$ when the emission of a Kelvin mode of positive amplitude slows the deepening rate back to what it was in the meridionally unbounded case. At $t = 20$ another Kelvin mode of positive amplitude is emitted and the deepening slows even more. At the eastern boundary, the raising of the height field behind the initial Rossby mode emitted at $t = 10$ proceeds as in the meridionally unbounded case until $t = 15$ when a Kelvin mode of negative amplitude arrives and is reflected as a set of Rossby modes of negative amplitude. The lowering of the

5. The height field at $t = 4$ at the west is not as deep as in the meridionally unbounded Figure 8a because the wind stress had to be turned on gradually over the first time period for numerical reasons.

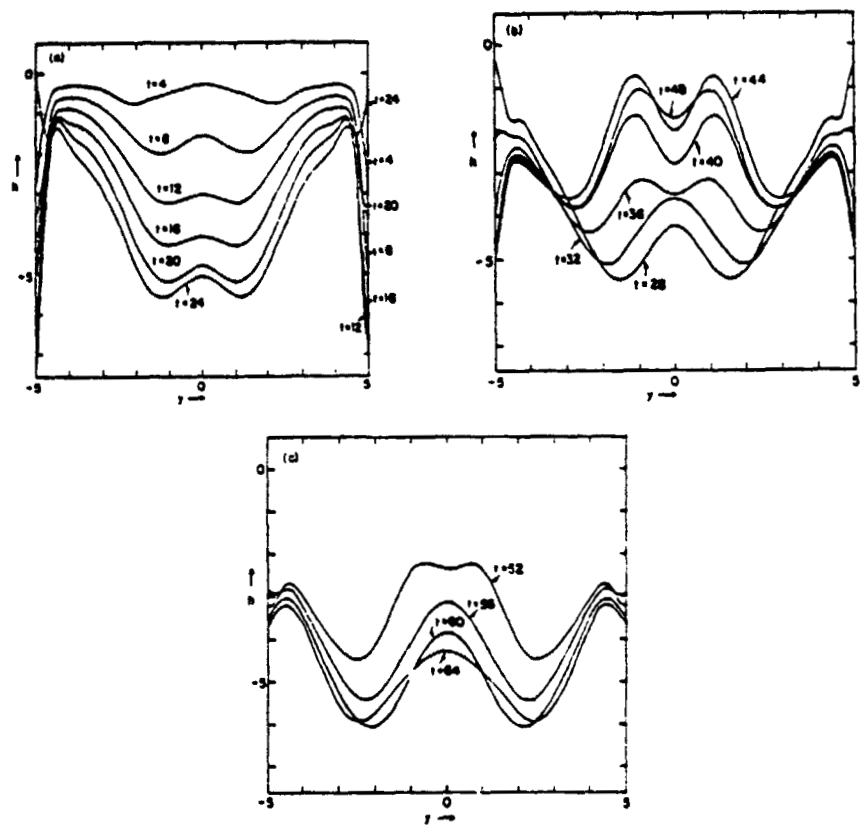


Figure 12. A meridional section of the height field at $x = 1$ from $[-5,5]$ for the $F = 1$ case. (a) Times $t = 4$ to 24; (b) Times $t = 28$ to 48; (c) Times $t = 52$ to 64.

height field between $t = 10$ and $t = 25$ therefore proceeds faster than in the meridionally unbounded case. Note in Figure 11a that despite the fact that the fronts have been smoothed by dispersion relative to Figure 8a, it is still possible to trace the Kelvin and the larger Rossby fronts across the basin.

At $t = 30$ the first $n = 1$ Rossby mode hits the western boundary emitting a positive amplitude Kelvin mode and at $t = 40$ the second negative amplitude Rossby mode hits the boundary emitting a negative Kelvin mode. The height field at the western boundary should then begin to increase at $t = 30$ and decrease at $t = 40$. Figure 11b, however, shows that the height field increases very slowly from $t = 32$ to 40 and continues to increase from $t = 40$ to $t = 48$ (it starts decreasing rapidly only after $t = 48$). What is happening is that wall-Kelvin waves hanging around from previous reflections are contaminating the results at the equator.

We can remove this contamination and see more clearly the progress of the response by examining a meridional slice of the height field at $x = 1$ (Fig. 12). As we

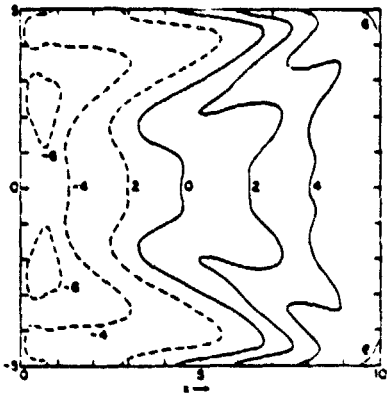


Figure 13. Same as Figure 10 but at $t = 64$.

have seen, the deepening proceeds rapidly from $t = 4$ to 20 (Fig. 12a). The slowing of the deepening at $t = 20$ and 24 due to the emission of the Kelvin modes at the western boundary at $t = 15$ and 20 is clearly seen as is the arrival of the anti-Kelvin modes at $t = 9$ and 14 which reverse the anti-Kelvin amplitude. The first Rossby hits $x = 1$ at $t = 29.7$ and by $t = 36$ (Fig. 12b) the characteristic $n = 1$ Rossby structure (compare to Fig. 5) has grown strong enough to be seen in the height field. The arrival of the strong anti-Kelvin at $t = 24$ begins to lower the anti-Kelvin amplitude as is clear in Figure 12b. The arrival of the second Rossby at $t = 39.7$ and its reflected negative amplitude Kelvin at $t = 41$ is most clear in the height field for $t = 52$ (Fig. 12c) where the large Rossby amplitude from the first (positive) Rossby mode has been effectively cancelled by the arrival of the second (negative) Rossby mode—only a remnant is left by $t = 52$. The arrival of the positive Rossby at $t = 59.7$ is slightly evident in the $t = 64$ curve.

A contour plot (Fig. 13) of the height field at $t = 64$ (the last time computed) shows that the height field is relatively well set up along the equatorial regions, less well set up away from the equator, and still less well set up away from the equator toward the west. This is understandable in terms of the large number of initial Rossby modes, $n = 1, 3, 5$, etc., that have reached longitudes close to the eastern boundary—at the other end of the basin, however, the $n = 3$ Rossby mode has not yet reached the western boundary by $t = 64$.

The gross energetics of the spin-up process are summarized by Figure 14a which shows potential and kinetic energy as a function of time. The kinetic energy is increasing rapidly with time until $t = 7.5$ when the initial Kelvin and Rossby modes meet and the initial rapid acceleration stops. The potential energy continues to increase as the initial Kelvin mode lowers the height at the west and the initial Rossby raises it at the east. It reaches a peak (i.e., the height field is maximally tilted) between $t = 16$ to 20 and then begins to decline as the emitted Kelvin modes in the

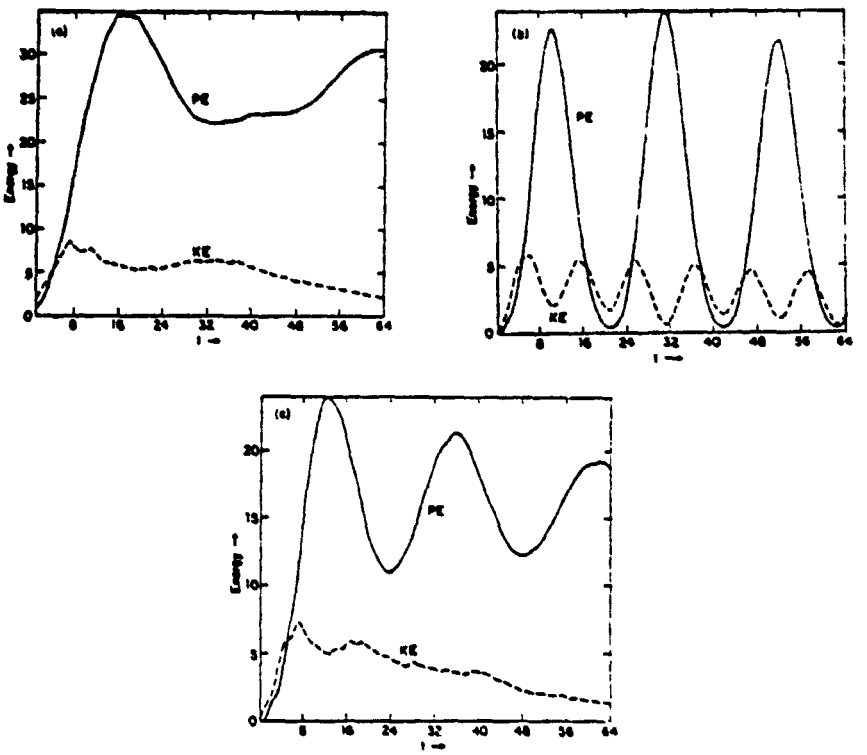


Figure 14. Energy versus time for the $F = 1$ case. (a) Basin [−5,5]; (b) Basin [−1.7,1.7]; (c) Basin [−5,1.7]. Units arbitrary.

west and the second Rossby mode in the east act to reduce the overall east-west tilt. The subsequent periodicity is approximately 40 time units—the roundtrip time of a Kelvin-Rossby mode transit.

The Rossby modes lose their importance in the [−1.7,1.7] case as can be seen from Tables 2 and 3. We may thus understand this case almost entirely in terms of the Kelvin and anti-Kelvin modes (since the northern and southern boundaries are so close to the equator there is considerable overlap between the Kelvin and anti-Kelvin modes and the delay due to the travel of the wall-Kelvin is substantially absent). The initial Kelvin mode emitted from the east at $t = 0$ has an amplitude of −2.36, carries all the mass flux of the meridionally unbounded solution (−2.88), and lowers the height field behind it as it propagates toward the east. The initial anti-Kelvin emitted from the east has amplitude −2.35, carries mass flux −2.76, and *raises* the height field as it propagates westward. Since both the Kelvin and anti-Kelvin work to increase the east-west tilt, the potential energy will increase monotonically until $t = 10$ when the Kelvin mode hits the western boundary. Since the anti-Kelvin mode carried almost all the mass flux, it reflects as a Kelvin mode of

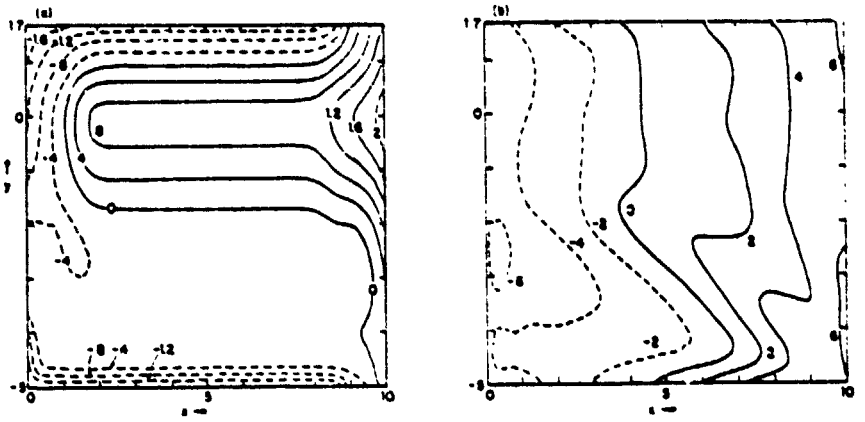


Figure 15. Same as Figure 10 but for a [−5,1.7] basin. (a) $t = 2$; (b) $t = 64$.

amplitude +2.26 which tends to raise the height field behind it. Similarly, the Kelvin mode reflects at the eastern boundary as an anti-Kelvin mode of amplitude +2.27, which now tends to lower the height field behind it. By $t = 20$ almost all the tilt induced by the initial Kelvin-anti-Kelvin pair is wiped out by the reflected anti-Kelvin-Kelvin pair: the height field becomes almost flat. The potential energy therefore has a periodicity of 20 time units and it almost reaches zero at its minimum; Figure 14b. The kinetic energy is largest when the height field is changing most rapidly and so has a periodicity half that of the potential energy.

The asymmetric basin [−5,1.7] provides an intermediate case, containing some features of both the wide basin and narrow basin case, as well as some features uniquely its own. Because of the lack of symmetry with respect to the equator, *all* modes will be excited. Tables 2 and 3 show that the $n = 0$ Rossby-Kelvin mode (Fig. 4b) is especially important in this asymmetric basin. Figure 15a shows the height field at $t = 2$. Now the amplitude on the northern wall is mostly Rossby-Kelvin while the amplitude on the southern wall is predominately anti-Kelvin. The dominant periodicity is now $t = 20$, the Rossby-Kelvin-Kelvin roundtrip transit time but, as can be seen in Figure 14c, the successive peaks are delayed because of the transit time of the wall-Kelvins needed to communicate between the Kelvin and anti-Kelvin modes. The state at the end of 64 time units, Fig. 15b, shows the height field evenly tilted in the northern part of the basin but again less well spun up in those parts of the southern region where the higher Rossby modes have yet to reach. Even in the northern regions, however, the height field continues to undergo oscillations as the Kelvin mode sloshes mass across the basin.

b. The $G = 1$ case

In the meridionally unbounded situation (discussed in Paper II), the $G = 1$ forcing excites no Kelvin or anti-Kelvin modes; only even n Rossby modes. By sym-

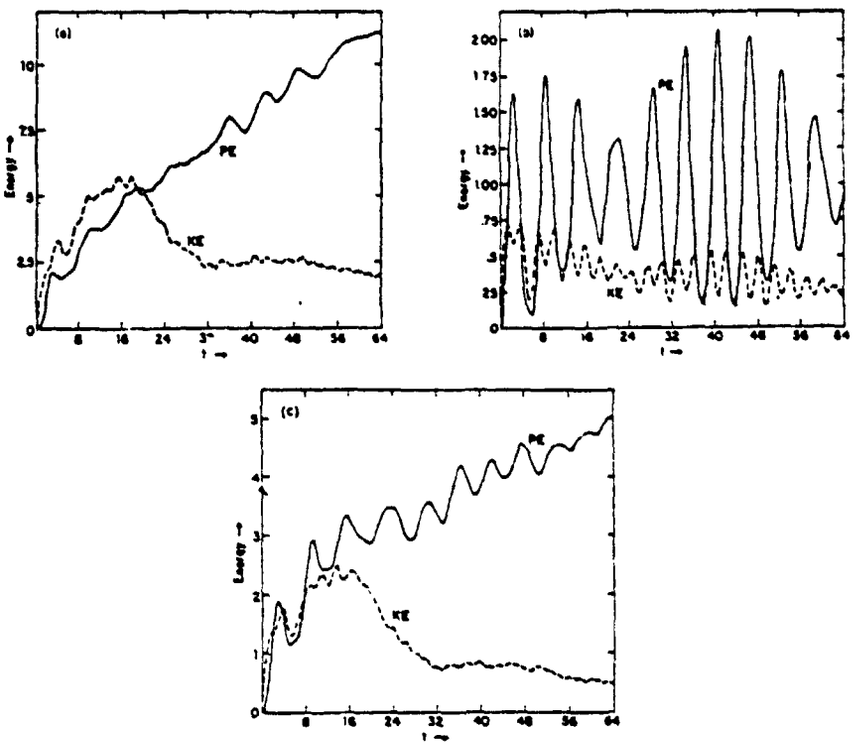


Figure 16. Energy versus time for the $G = 1$ case. (a) Basin [−5,5]; (b) Basin [−1.7,1.7]; (c) Basin [−5,1.7].

metry considerations, none of these modes contains any net meridionally integrated zonal mass flux, so that upon reaching the western boundary no Kelvin waves will ever be excited. According to Eq. (55), the zonally unbounded solution plus all its reflections sum simply to $(0,0,y)$, the final steady state. The response therefore proceeds relatively straightforwardly: the more Rossby waves emitted from the eastern boundary at $t = 0$ that reach a given point, the closer is the height field at the point to the final state. Thus points closer to the equator and to the eastern boundary spin up faster. The spin-up is monotonic with time—no Kelvin waves exist to slosh mass across the basin.

The response in the [−5,5] basin is very much the same as in the meridionally unbounded case except that the lowest planetary mode excited at the eastern boundary is now the $n = 0$ Rossby-Kelvin mode which does not exist in the meridionally unbounded case. Figure 16a shows the energy diagram for the [−5,5] basin. The wiggles, of period about 2π, are the initial $n = 0$ inertia-gravity waves (at $k = 0$, $\omega \cong 1$ in Fig. 1) which are excited initially to produce the zonally unbounded solution $(u^{(2)}, 0, h^{(2)})$ in Fig. 6a. Because the zonally unbounded state has a height field

which is tilted near the equator some moderate fraction of the final height field tilt and because this zonally unbounded state does not grow with time, the magnitude of the intertia-gravity oscillations are moderate and remain of constant amplitude. The initial $k = 0$, $\omega \cong 1$ has a group velocity to the east and reflects as a $k = -1$, $\omega \cong 1$ inertia-Kelvin wave—these reflections continue forever superimposed on the monotonically spinning-up height field.

Meridional sections of the height field in the [−5,5] case taken toward the end of the simulation ($t = 64$) indicate that the eastern part of the basin is almost completely spun up, as we expect, while the degree of spin-up at $t = 64$ decreases as we move westward across the basin. Because the $n = 0$ mode has its largest amplitude at the northern and southern walls (Fig. 4a) and travels with speed near unity, the regions near these walls are spun-up even at the west while substantial parts of the region between these walls and the equator have not yet felt the influence of any Rossby modes from the east and so are far from spun-up.

The energetics of the response in the [−1.7,1.7] case are shown in Figure 16b. In this case the inertia-gravity waves needed to set up the zonally unbounded solution have extremely large amplitude because the height field of this zonally unbounded solution (Fig. 6c) has a very substantial part of the tilt of the final spun-up height field. The envelope of the potential energy is modulated by the $n = 0$ inertia-gravity wave roundtrip transit time (calculated from the group velocities) which in the [−1.7,1.7] case is about 40 time units.

The asymmetric basin case, [−5,1.7], is completely different from the symmetric case in that the asymmetry of the basin modes now implies that *all* the modes carry net meridionally integrated zonal mass flux. Thus, in accordance with our discussion of western boundary response, Kelvin waves will be generated when any of the westward propagating modes reach the western boundary. Examining Table 2, we see that the zonally unbounded solution ($u^{(2)}$, 0, $h^{(2)}$) plus the reflections at the eastern boundary lead to a height field $h = y + .0986$ at the eastern boundary (by Eq. (55)), and by extension, to any point in the interior to which all the Rossby modes produced by this reflection have reached. The final steady state height field, however, is $h = y + 1.65$ so that even after all the Rossby modes due to the first reflection have reached all points, the tilt would be correct but there would still be a height deficit of 1.55 units throughout the basin. The additional mass, of course, is carried by the Kelvin wave which must reflect the total mass flux of the unbounded solution, −1.596. This Kelvin wave, of amplitude $b_E = 1.250$, thus carries a mass flux +1.596 and when at $t = 10$ (plus the wall-Kelvin travel time) this Kelvin mode hits the eastern boundary, it, plus its reflections, uniformly raises the height everywhere along the eastern boundary by an amount .979 according to Eq. (52) and Table 3. The initial Kelvin wave alone thus fills two-thirds of the height deficit. The initial anti-Kelvin, however, upon reflection at the western boundary, produces a Kelvin of negative amplitude −.786 which, when it hits the eastern boundary, lowers the

height field uniformly at the east by an amount .616, undoing some of the good works of the first Kelvin. From this point on, most of the mass flux in the succeeding reflections is sloshed around by the Kelvin mode reflecting into a Rossby-Kelvin mode and vice versa with a periodicity of 20 time units. Figure 16c shows the energy diagram for this case and we see that a plateau exists from $t = 20$ to 40 but that the inertia-gravity waves tend to mask the effect.

In order to filter out these inertia-gravity waves, we reran the $G = 1$, $[-5,1.7]$ case starting, not from a flat ocean at rest, but rather from a state from which the inertia-gravity waves had already been removed, namely, the $(u^{(2)}, 0, h^{(2)})$ state. This filtering was largely successful and a comparable energy diagram shows that the 20 unit Rossby-Kelvin-Kelvin periodicity is clearly apparent. As time goes on, the height field would tilt closer to its final state, but contrary to the symmetric basin $G = 1$ cases, mass would continue to be sloshed across the basin.

6. Conclusion

This paper has presented a straightforward analytic approach to the calculation of the time-dependent response of equatorial basins to wind stresses varying slowly in the zonal direction and in time.[6] This allows (to our knowledge, for the first time) a rather detailed comparison of analytic theory with linear simulations, or equivalently, with the early (linear) stages of nonlinear simulations. The method involves first approximating and describing the free modes of a meridionally bounded basin; next calculating the meridionally bounded but zonally unbounded solution which gets set up by the emission of inertia-gravity waves within a single inertial period; then calculating the eastern and western boundary responses to this zonally unbounded solution and to any subsequent responses; and finally, following in time the series of reflections and responses that leads, or does not lead, to the steady state solution. The power of the method to accurately describe linear spin-up was illustrated by a series of numerical simulations of the responses of three separate ocean basins, $[-5,5]$, $[-5,1.7]$ and $[-1.7,1.7]$, all 10 units long in the zonal direction, to the simple wind stress forcings $F = 1$ and $G = 1$, and following the response out to 64 nondimensional time units for each case.

Compared to the meridionally unbounded case treated in Paper II, we saw that the introduction of the northern and southern boundaries has in some ways simplified the consideration of spin-up and in some ways complicated it. For example, the problem is simpler in the $F = 1$ symmetric basin cases compared to the meridionally unbounded case in that only those modes whose turning points lie within the basin carry any meridionally integrated zonal mass flux, and all those modes whose turning points would have lain outside the basin in the meridionally unbounded case and which would have been needed to account for the rearrangement of mass dur-

6. Only winds independent of x and switched on in t were calculated explicitly but the extension to winds slowly varying in x and t is straightforward using the methods of Paper I.

ing spin-up are, in the symmetric basin case, replaced by a single anti-Kelvin mode. On the other hand, the problem is more complicated in these same cases in that the additional reflections induced by this same anti-Kelvin mode lead to a much more intricate reflection diagram (see Fig. 9). In the narrow basin case, [−1.7,1.7], the description becomes simpler still: since all Rossby modes higher than $n = 1$ have their turning points outside the basin, the complete system can be described by only three modes—the Kelvin, the anti-Kelvin and the $n = 1$ Rossby mode.

The results of the theory presented in Sections 2, 3, and 4, and of the simulations analyzed in Section 5, now allow us to address a question raised in the Introduction: to what extent does the imposition of the northern and southern boundaries affect the progress of the time-dependent response in the vicinity of the equator? The answer, as we saw, varies from case to case. The case in which the boundaries produced the least effect was the $G = 1$, [−5,5], case. Here the spin-up in the region of the equator was precisely the same as in the meridionally unbounded case. The only difference was that the fast $n = 0$ Rossby-Kelvin mode, with large amplitude at the northern and southern boundaries, spun up the regions near these boundaries rapidly, whereas this mode is totally absent in the unbounded case. In the $G = 1$, [−1.7,1.7], case, the boundaries are so close to the equator that the Rossby-Kelvin mode has considerable amplitude even at the equator. The entire basin would therefore spin-up relatively completely after only 10 time units (one Rossby-Kelvin mode traversal time) were it not for the large amplitude inertia-gravity waves set up during the initial unit of time. We can conclude that for $G = 1$, in basins symmetric with respect to the equator, the time-dependent response near the equator will proceed essentially unaffected by the northern and southern boundaries as long as these boundaries lie at least a few equatorial radii of deformation from the equator. For the $G = 1$ case, in basins asymmetric with respect to the equator, or for the $F = 1$ case in any of the basins, the extra Kelvin modes induced by the western boundary reflections of either anti-Kelvin or Rossby-Kelvin modes *will* change the details of the time-dependent response near the equator—the closer a boundary is to the equator the larger will be the change.

The role of the Kelvin mode in the spin-up process bears repeated comment. The *initial* Kelvin mode, in all those cases in which it exists ($G = 1$, asymmetric basin, and $F = 1$, all basins) helps to bring the height field closer to its spun-up (steady) value. On the other hand, repeated production of Kelvin modes by reflection of any of the westward propagating modes induces a sloshing of mass back and forth across the basin which inhibits the attainment of the final spun-up state. In the absence of friction, this sloshing would continue forever. Only in those cases in which the Kelvin mode never appears ($G = 1$, symmetric basins) does the time-dependent response truly proceed to a steady state.

It is clear that the method and results of this paper (and Papers I and II) are directly useful in analyzing numerical simulations of time-dependent response. But

in application to the real ocean we must offer a major *caveat*. We have assumed the existence of a single baroclinic mode. While there is evidence that standing vertical modes do exist (Wunsch and Gill, 1976), the question has recently been raised (Philander, 1978) of whether or not standing vertical modes are regularly excited by atmospheric forcing. Since the dynamics of baroclinic mode excitation is not at all well understood, we would urge the greatest caution in applying this, or any theory based on a single standing vertical mode, to the behavior of real oceans.

Acknowledgments. We would like to thank Brian Hickie for making available a copy of his paper prior to publication; Moshe Israeli for valuable help with the numerical work; Richard Lindzen for discussions of positive equivalent depth modes on β-planes; Dennis Moore and George Philander for continued discussions about the dynamics of equatorial waves; and D. R. Sadigur for careful preparation of the manuscript. This work was supported at Harvard University by NASA Grant NSG-5160.

APPENDIX: Orthogonality and completeness of the eigenfunctions

Fourier transforming the unforced ($F = G = Q = 0$) Eqs. (1) from x to k and denoting transpose by superscript T we may write (cf. I (36)ff):

$$\frac{\partial}{\partial t}(u,v,h)^T + i\,\mathrm{A}(u,v,h)^T = 0$$

where

$$\mathrm{A} \equiv \begin{bmatrix} 0 & iy & k \\ -iy & 0 & l\frac{\partial}{\partial y} \\ k & l\frac{\partial}{\partial y} & 0 \end{bmatrix}$$

then $\mathrm{A}\,\Phi_{n,l} = \omega_{n,l}\,\Phi_{n,l}$ and the following five lemmas follow easily:

Lemma 1 $\mathrm{A} = \mathrm{A}^*$ [Readily shown by direct calculation.]

Lemma 2 All the ω's are real. [Follows from the self-adjointness of A, that is, from Lemma 1.]

Lemma 3 For all n, $\mu_n > 0$ [Since (1) and (2) form a standard Sturm-Liouville problem.]

Lemma 4 The ω's are distinct. [Follows from Lemma 3, the dispersion relation and a consideration of the Kelvin roots $\omega = \pm k$.]

Lemma 5 The $\{\Phi_{n,l}\}$ is orthogonal. [Follows from Lemma 4 in the usual way, i.e.

$$\omega_{m,l}\,[\Phi_{n,\mu}, \Phi_{m,l}] = [\Phi_{n,\mu}, \mathrm{A}\,\Phi_{m,l}] = [\mathrm{A}\,\Phi_{n,\mu}, \Phi_{m,l}] = \omega_{n,\mu}\,[\Phi_{n,\mu}, \Phi_{m,l}]$$

with the second equality a consequence of Lemma 1. Since the ω's are distinct, the Φ's are orthogonal.]

Theorem: $\{\Phi_{n,l}\}$ is complete.

Proof:

First rephrase the question of completeness as follows. Let $R = \{\Phi_{n,l} | n \geq 0\}$, i.e., all the eigenvectors except the Kelvin waves $\phi_{K,m}$. We wish to show that if

$$[(U, -iV, H)^*, \Phi_{n,l}] = 0 \text{ for all } \Phi_{n,l} \in R$$

then

$$V = 0;\ yH + U' = 0;\ \text{and } yU + H' = 0. \tag{A.1}$$

Since the only solutions to (A.1) are (multiples of) $\Phi_{K,m}$ the theorem will follow.

It is straightforward to show from the definition (37) of $\Phi_{n,j}$ that

$$[(U, -iV, H)^*, \Phi_{n,j}] = N_{n,j}^{-1} [A_{n,j}, \psi_n]$$

where

$$A_{n,j} = \omega_{n,j} (yU + H') + k(yH + U') + (\omega_{n,j}^2 - k^2) V .$$

Now by assumption

$$0 = [A_{n,j}, \psi_n] \text{ for all } j, n \geq 0 . \tag{A.2}$$

Hence

$$0 = [A_{n,j} - A_{n,l}, \psi_n] = (\omega_{n,j} - \omega_{n,l})[yU + H' + (\omega_{n,j} + \omega_{n,l}) V, \psi_n]$$

since

$$\omega_{n,j} \neq \omega_{n,l} \text{ for } j \neq l$$

$$(yU + H' + [\omega_{n,j} + \omega_{n,l}] V, \psi_n) = 0 \text{ for all } n \geq 0, j \neq l . \tag{A.3}$$

Let $l = 3$. Then taking the difference of (A.3) with $j = 1$ and $j = 2$ and using $\omega_{n,1} \neq \omega_{n,2}$ yields

$$(V, \psi_n) = 0 \text{ for all } n . \tag{A.4}$$

Since the ψ_n's are complete, $V = 0$. Then (A.3) implies that $yU + H' = 0$ and, finally, (A.2) yields $yH + U' = 0$, completing the proof.

In the body of this paper we make particular use of the completeness when $k = 0$. In this case the proof must be modified slightly. It is unchanged through (A.4) after which we must use

$$[(U, iV, H)^*, \Phi_{n,0}] = -(2\mu_n + 1)^{-1} (yU + H') + (yH + U')$$

to establish that

$$yH + U' = 0 .$$

REFERENCES

Abramowitz, M. and I. A. Stegun. 1965. Handbook of Mathematical Functions, New York, Dover, 1046 pp.

Anderson, D. L. T. and P. B. Rowlands. 1976. The role of inertia-gravity and planetary waves in the responses of a tropical ocean to the incidence of an equatorial Kelvin wave on a meridional boundary, J. Mar. Res., *34*, 295–312.

Cane, M. A. 1975. A study of the wind-driven ocean circulation in an equatorial basin. Ph.D. Thesis, Massachusetts Institute of Technology.

Cane, M. A. 1979. The response of an equatorial ocean to simple wind stress patterns: I. Model formulation and analytic results, J. Mar. Res., *37*, 233–252.

Cane, M. A. and E. S. Sarachik. 1976. Forced baroclinic ocean motions. I. The linear equatorial unbounded case. J. Mar. Res., *34*, 629–665.

Cane, M. A. and E. S. Sarachik. 1977. Forced baroclinic ocean motions. II. The linear equatorial bounded case. J. Mar. Res., *35*, 395–432.

Hickie, B. P. B. 1979. The effects of coastal geometry on equatorially trapped planetary waves. Part I: Free oscillations in the Gulf of Guinea. To appear.

Matsuno, T. 1966. Quasi-geostrophic motions in the equatorial area. J. Met. Soc. Japan, *44*, 25–43.

Mofjeld, H. O. and M. Rattray, Jr. 1971. Free oscillations in a beta-plane ocean. J. Mar. Res., *29*, 281–305.

Moore, D. W. 1968. Planetary-gravity waves in an equatorial ocean. Ph.D. Thesis, Harvard University.

Moore, D. W. and S. G. H. Philander. 1976. Modelling of the tropical oceanic circulation, *in* The Sea, Vol. 6, Chapter 8, Goldberg *et al.*, eds., New York, Interscience, 1048 pp

Philander, S. G. H. 1978. Forced oceanic waves. Revs. Geophys. Space Phys., *16*, 15–46.

Wunsch, C. and A. E. Gill. 1976. Observations of equatorially trapped waves in Pacific sea level variations. Deep Sea Res., *23*, 371–390.

Received: *18 August, 1978;* revised: *3 March, 1979.*

Printed in U.S.A. for the Sears Foundation for Marine Research,
Yale University, New Haven, Connecticut, 06520, U.S.A.
Van Dyck Printing Company, North Haven, Connecticut, 06473, U.S.A.

Paper 10

Reprinted from
(COSPAR) REMOTE SOUNDING OF THE ATMOSPHERE FROM SPACE
Edited by
H. J. BOLLE

PERGAMON PRESS OXFORD and NEW YORK 1979

EFFECTS OF SOUNDING TEMPERATURE ASSIMILATION ON WEATHER FORECASTING: MODEL DEPENDENCE STUDIES

M. Ghil*, M. Halem** and R. Atlas***

*Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, and Goddard Laboratory for Atmospheric Sciences
**Laboratory for Atmospheric Sciences, NASA Goddard Space Flight Center, Greenbelt, MD 20071
***Department of Mechanical Engineering, State University of New York at Stony Brook, NY 11794, and Goddard Laboratory for Atmospheric Sciences

INTRODUCTION

In comparing various methods for the assimilation of remote sounding information into numerical weather prediction (NWP) models, the problem of model dependance for the different results obtained is clearly important. Most assimilation experiments have not addressed this problem directly, every research and development group working essentially with one numerical model.

We have started to investigate two aspects of the model dependence question: 1) the effect of increasing horizontal resolution within a given model on the assimilation of sounding data, and 2) the effect of using two entirely different models with the same assimilation method and sounding data. In order to study the first aspect, we have applied our four-dimensional (4-D) time-continuous statistical assimilation method (SAM: [1,2,3,4]) to the assimilation of sounding data from the Data Systems Test (DST) period DST-6 (February 1976) into the Goddard Laboratory for Atmospheric Sciences (GLAS) general circulation model (GCM) with a coarse resolution (4° lat.x5° long.) and with a fine resolution (2.5° lat.x3° long.). To study the second aspect of model dependence, we have applied an asynoptic, time-continuous successive correction method (SCM: [2,3,5]) to the global, nine-level, primitive-equation model of the U.S. National Meteorological Center (NMC: [6]). Results of SCM assimilations with the latter model are compared with those of previous SCM experiments with DST-6 data using the GLAS coarse-mesh model.

DEPENDENCE ON MODEL RESOLUTION

Preliminary results of comparisons between SAM assimilations with the coarse-mesh and the fine-mesh GLAS model show that increased model resolution does not affect the effect of satellite data on weather forecasts in any negative way; the effects of increased resolution and of satellite data assimilation are cumulative: they both tend to improve forecasting accuracy by a similar fractional amount. This fact is reflected in the usual numerical measures of forecast accuracy, viz., RMS errors and S_1 skill scores. It is also reflected in synoptic case studies of the large-scale forecast fields.

Effect on Numerical Measures of Forecast Accuracy

Assimilation cycles were run using data from 29 January till 21 February 1976 with both the coarse and the fine GLAS models. For each model resolution, one cycle was run including all data available during the DST-6 period, in particular temperature sounding data from the NOAA-4 and Nimbus-6 satellites, and another one was run with the same data, but excluding the temperature sounding information: they are referred to respectively as the 2SAT and the NOSAT cycle. For each one of these four experiments, 72h forecasts were performed from the analyses produced by the corresponding assimilation cycle at 48h intervals, starting on 1 February. The forecasts were verified at 48h and at 72h against NMC operational analyses. The quantities verified and regions of verification are the same as in [1,2,3,4,5].

The results are summarized in Table 1, according to the methods outlined in [1,2,3, 4,5]. It is seen that the effect of model improvement, specifically of increased resolution on the one hand, and of utilization of satellite data on the other, act both in the direction of improving the forecast accuracy. In fact, the two effects appear to be nearly additive, as indicated by a comparison of the results in the last three rows of the table. Note also that the 2SAT experiment 2F, with the fine model, consistently gives impacts which are significantly higher than those of the corresponding experiment with the coarse model,2C, when measured against the common standard of the coarse NOSAT experiment 0C (viz. the second and the last row of the table).

TABLE 1 Summary of Results for Model Dependence Experiments

Experiment No.	Comparison Code	Impact, % S_1	Impact, % RMS	Statistical Significance (average/standard error) S_1	Statistical Significance (average/standard error) RMS
0C	0C	0	0	-	-
2C	2CC	4.46	12.41	1.85	2.59
0F	0F	0	0	-	-
	0FC	5.86	7.30	1.62	1.35
2F	2FF	4.49	7.98	1.40	1.71
	2FC	10.01	14.47	2.67	2.47

The symbol 0 stands for a NOSAT experiment, 2 stands for a 2SAT experiment; the symbol C stands for an experiment with the coarse model, F for one with the fine model. The results of the 2SAT experiment with the fine moel, 2F, were compared both with those of the coarse NOSAT experiment 0C, and with those of the fine NOSAT experiment 0F; these comparisons are denoted by 2FC and by 2FF, respectively. The NOSAT experiment 0F appears both as a control for 2F, viz., as 0F, and in a comparison versus the coarse NOSAT 0C, viz., as 0FC.

Effect on Synoptically Significant Forecast Improvements

It is well known that moderate improvements of forecast accuracy, as measured by RMS errors and by S_1 skill scores, are not consistently reflected in the improved prognosis of major synoptic features. Within a certain number of forecasts which

show a slight improvement in their statistics, only a small fraction will show actual changes for the better which are noticeable synoptically. This was observed also in the experiments reported herein.

A preliminary examination of the numerically predicted flow patterns showed no major changes for the worse in any 2SAT forecast with the fine model, as compared with the corresponding NOSAT forecast, minor changes for the better which slightly outweigh those for the worse, and at least one major improvement. This noticeable improvement attributable to proper utilization of satellite data occurred in the 72h forecast from 0000GMT 19 February 1976.

At the initial time of the forecast, the analysis showed a moderately intense low-pressure system located off the northwest coast of the United States. As this system moved inland, a new low developed along an already existing stationary front and became the dominant feature. During the first 48 hours of the model forecast only slight differences exist between the 2SAT and NOSAT sea-level prognoses as both correctly predict the intensification and southeastward displacement of this low. After 48 hours however a radically different pattern emerges: the 2SAT forecast predicts the low to recurve and accelerate to the northeast, while the NOSAT forecast fails to move the low center. The NMC operational PE hemispheric forecast model also failed to move the low northeastward. This low was associated with both heavy snow and severe local storms during its displacement, and significantly different local weather forecasts of those major events would have resulted from the differing prognoses.

Figures 1a and 1b depict the 72h sea-level pressure prognoses from 19 February for the 2SAT and NOSAT cases respectively, while the corresponding analysis is depicted in Fig. 1c. Figure 1d shows the NMC operational 72h forecast. A comparison of these charts reveals the very significant improvement in the displacement of the low-pressure system that resulted from the inclusion of satellite data. Substantial improvements in the displacement of the associated upper-level trough and vorticity maximum (charts not shown) were also observed.

INTER-MODEL DEPENDENCE

Previous results with the GLAS coarse-mesh model [1,2,3,4,5] indicated that by using a time-continuous 4-D method with statistical features (SAM or SCM) the assimilation of remote sounding temperature data can lead to numerical forecasts which show a modest, but statistically significant improvement over forecasts started from analyzed fields not containing the remote sounding information. To determine whether the translation of our methods into operational practice is feasible, we decided to apply them to the model used by NMC to produce their operational analyses [6], which shall be called in the sequel the 9LG (nine-level global) model.

As a first step, SCM was adapted to and reprogrammed for the 9LG model. Numerous problems arose in the adaptation, due to differences between the models and some of these problems await better solutions than the ones given to them at the present. A NOSAT and a 2SAT assimilation cycle were started and carried out from 1 February till 7 February 1976, using the same data sets as for the corresponding GLAS model cycles. Forecasts were carried out from the analyzed fields generated on 3, 5 and 7 February. These experiments are being continued and will be discussed more completely in a future publication.

At this point, a comparison of results with those of SCM experiments performed on the GLAS model [2,3,4,5] shows that the differences between analyzed fields from the NOSAT and 2SAT cycles do not seem to depend on the model: the differences produced by our utilization method of satellite data are of similar magnitude for the 9LG model as they were for the GLAS 4°x5° model. These differences range up

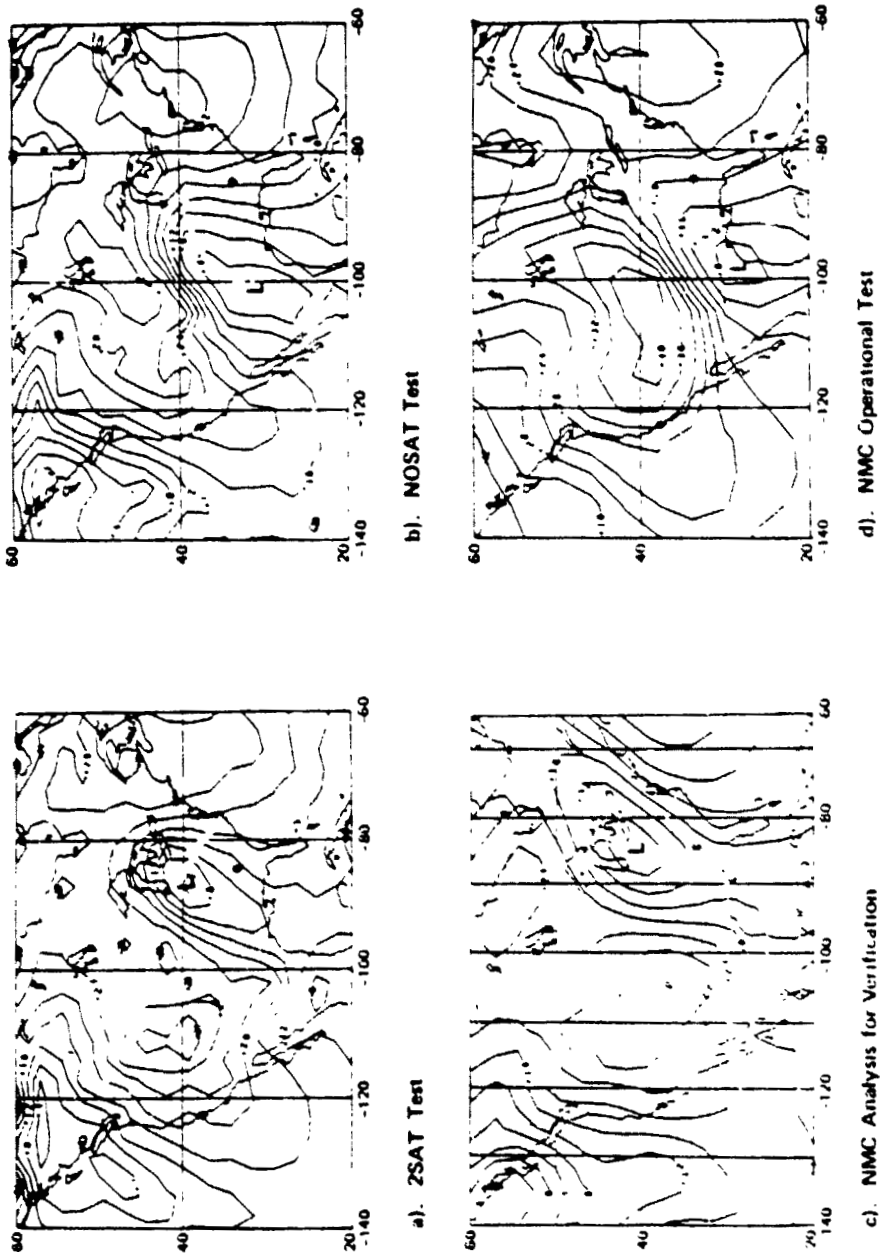


Fig 1. Sea-level pressure plots for 0000GMT 22 February 1976. Figure 1a shows the 72 h forecast from 2SAT initial data at 0000GMT 19 February 1976; Figure 1b the corresponding NOSAT forecast; Figure 1c shows the verification analysis (operational NMC analysis), and Figure 1d the NMC operational 72 h forecast. Notice the close similarity between 1a and 1c in placing the low-pressure system over the Ohio Valley, while 1b and 1d place it over Western Texas.

to 100 m in the 500 mb geopotential height field, especially over the North Pacific, where conventional data coverage is relatively poor. These initial state differences lead also to correspondingly large differences in 48h and 72h forecasts. We intend to examine further the question whether these differences are due solely to the assimilation of satellite data.

CONCLUDING REMARKS

Our experiments on model dependence of sounding temperature assimilation suggest the following tentative conclusions: First, that model improvement, as exemplified by increased resolution, can act in the same direction as judicious 4-D assimilation of remote sounding information, to improve 2-3 day numerical weather forecasts. The effect of model improvements and the effect of an observational data base augmented by the proper inclusion of sounding temperatures, when combined, can lead to numerical prognoses which achieve improvements of synoptical significance more easily than either effect separately. Second, that the time-continuous 4-D methods developed at GLAS have similar beneficial effects when used in the assimilation of remote sounding information into NWP models with very different numerical and physical characteristics. This suggests the possibility of the eventual implementation of these time-continuous methods in an operational environment, in which they might contribute to noticeable improvement in routine weather forecasting.

REFERENCES

1. M. Ghil, R. Dilling and H. Carus, in: Proc. 5th Conf. Probability Statistics Atmos. Sci., Amer. Meteor. Soc., Boston, MA, 1977, 320-324.

2. _____, M. Halem and R. Atlas, Mon. Wea. Rev., in press.

3. M. Halem, M. Ghil, R. Atlas, J. Susskind and W. J. Quirk. The GISS Sounding Temperature Impact Test. NASA Tech. Memo. 78063, Goddard Space Flight Center, Greenbelt, MD, 1978, 421 + xiii pp.

4. _____, M. Ghil and R. Atlas, in: Remote Sensing of the Atmosphere: Inversion Methods and Applications, A. L. Fymat and V. E. Zuev, eds., Elsevier, Amsterdam, 1978.

5. M. Ghil and R. Dilling, in: NASA TM 78063, 1978, pp. 3.1-3.173.

6. J. D. Stackpole, in: Proc. 6th Conf. Weather Forecasting Analysis, Amer. Meteor. Soc., Boston, MA, 1976, 112-116.

Paper 11

Reprinted from *Monthly Weather Review*, 107, pp. 140-171, 1979.

Time-Continuous Assimilation of Remote-Sounding Data and Its Effect on Weather Forecasting

M. GHIL[1], M. HALEM AND R. ATLAS

Laboratory for Atmospheric Sciences, NASA Goddard Space Flight Center, Greenbelt, MD 20771

(Manuscript received 7 April 1978, in final form 6 October 1978)

ABSTRACT

Methods are derived for the time-continuous four-dimensional assimilation of satellite sounding temperatures. The methods presented include time-continuous versions of direct insertion, successive correction and statistical linear regression. They are applied to temperature sounding data obtained from radiance measurements taken by instruments aboard the polar-orbiting satellites NOAA 4 and Nimbus 6. The data were collected during the U.S. Data System Test in January–March 1976.

A comprehensive series of experiments was performed to study the effects of using various amounts of satellite data and differing methods of assimilation. The experiments included the assimilation of data from the NOAA 4 satellite only, from Nimbus 6 only, and of data from both satellites combined. Other experiments involved variations in the application of our time-continuous statistical assimilation methods and of asynoptic successive correction methods. Intermittent assimilation of the sounding data was also tested, and its results compared with those of time-continuous assimilation.

Atmospheric states determined in the assimilation experiments served as initial states for a sequence of evenly spaced 3-day numerical weather forecasts corresponding to each experiment. The effects of the satellite data were evaluated according to the following criteria: 1) differences between the initial states produced with and without utilization of satellite data, 2) differences between numerical predictions made from these initial states, and 3) differences in local weather forecasts resulting from the large-scale numerical predictions.

Initial-state differences were evaluated in terms of magnitude and location of large-scale differences between meteorological fields. Numerical prediction differences were evaluated in terms of S_1 skill scores and rms errors, as well as by synoptic case studies. An automated forecasting model (AFM) based on quasi-geostrophic theory and on subjective forecasting principles was developed to facilitate the objective evaluation of differences produced in local weather forecasts, especially precipitation forecasts.

These studies suggest the following conclusions: 1) satellite-derived temperature data can have a modest, but statistically significant positive impact on numerical weather prediction in the 2–3 day range; 2) the impact is highly sensitive to the quantity of data available, and increases with data quantity; and 3) the method used to assimilate the satellite data can influence appreciably the magnitude of the impact obtained for the same data.

TABLE OF CONTENTS

1. Introduction

One of the main reasons we cannot tell the weather tomorrow is that we do not know the weather today. Numerical weather prediction (NWP) models require the complete specification of all the initial state variables of the atmosphere, yet only a fraction of the required initial data are available at any given time. The required data consist of the three-dimensional fields of temperature, humidity and horizontal velocity at locations comparable in number and spatial distribution with the number of grid points of the models used for NWP. By long-standing convention, such data are furnished at the so-called synoptic times, 0000 and 1200 GMT, and numerical forecasts are performed every day starting with data at these initial times.

The data available at synoptic times fall far short of being sufficient in number; also, in contradistinction to the regular spacing of model grid points,

[1] Courant Institute of Mathematical Sciences, and Laboratory for Atmospheric Sciences, New York University, New York, NY 10012.

0027-0644/79/020140-32$12.00

they are very unevenly distributed. Much of the data are found to be concentrated over the continents of the Northern Hemisphere, while the oceans and the Southern Hemisphere are regions where the data coverage is very sparse.

The advent of meteorological satellites in the late 1960's seemed to offer an effective way to supplement the conventional synoptic data network. Satellites carrying infrared and microwave scanning sounders offered the possibility of obtaining vertical temperature profiles with global coverage by remote-sounding techniques.

Satellite-sounding data, though promising, provide only one of the basic variables, viz., temperature. Moreover, while the conventional data sets are available simultaneously over the entire globe, temperature sounding data are obtained only in asynoptic, time-continuous fashion. In one of the earliest efforts at using remote-sounding temperatures for NWP, Charney *et al.*, (1969) put forward the conjecture that a complete knowledge of the continuous temperature history of the atmosphere will determine other initial state variables, in particular the winds. A great number of numerical experiments with general circulation models were carried out based on this conjecture. Useful reviews of these experiments can be found in Bengtsson (1975), Jastrow and Halem (1973), Kasahara (1972) and McPherson (1975).

Recently, Ghil *et al.* (1977a) have given a precise formulation to the Charney conjecture for certain simple atmospheric models; they proved it analytically to be correct for these models within the formulation given. However, in practice, numerous problems with real data and with complexities of current NWP models are not completely covered by their rigorous theory. Therefore, it seems doubtful whether a constructive method can be devised, which from the continuous history of the temperature field alone will yield initial states of arbitrary accuracy.

The use of temperature history to obtain initial states has become known in the literature as four-dimensional (4D) assimilation of temperatures. In the present article we compare a number of methods of 4D assimilation of remote-sounding temperatures and their effect on the accuracy of the initial states obtained, as well as on the resulting forecasts generated from these initial states.

The methods considered herein are mostly time-continuous; one assimilation experiment using intermittent assimilation of asynoptic data at synoptic times was carried out for the purpose of comparing the present methods with more conventional data analysis techniques. The methods presented include direct insertion (DIM), successive corrections (SCM) and a statistical assimilation method (SAM).

In addition to the methods mentioned above, a number of other methods, whose development or testing is not complete at present, are reported in Halem *et al.* (1978). These include an asynoptic variational method using a direct minimization technique (Ghil and Mosebach, 1978) and assimilation using a set of filtered equations (Peng and Shkoller, 1978).

Assessment of the effect that sounding data and methods for their assimilation have on weather forecasting is complicated by the verification problem. This problem has to do with a lack of consensus on adequate measures for the accuracy and utility of a large-scale numerical forecast. In the present study a new verification criterion is introduced based on the accuracy of local precipitation forecasts derived from large-scale numerical model forecasts. An automated forecasting model (AFM) is formulated to allow for the rapid computation of large numbers of local weather forecasts. The AFM is then applied to the evaluation of our experiments, along with the usual S_1 and rms error measures, and with a number of subjective case studies.

All the experiments described herein were performed on a standard set of data: those provided by the National Meteorological Center (NMC) and by the National Environment Satellite Service (NESS) during the Data System Test period DST-6 (January–March 1976). Experiments were carried out to study the effect of data quantity, as well as the role of assimilation method.

The assimilation and verification methodology is developed in Section 2. The numerical experiments are described in Section 3. Their results are presented and discussed in Section 4. Conclusions follow in Section 5.

2. Description of the methods

The main contribution of this article consists in the way data are assimilated and in some of the criteria used for assessing the impact. The assimilation methods considered fall into the broad category of time-continuous methods. Time-continuous methods are particularly appropriate for the assimilation of remote-sounding temperature data because of the asynoptic nature of these data. In the following we shall present a number of time-continuous versions of direct insertion, of successive correction techniques and of statistical assimilation methods. The latter methods are currently gaining greater recognition and wider acceptance for the purpose of assimilating different types of observational data in the process of determining the initial states from which numerical forecasts are made (Phillips, 1976).

All the methods described were applied to assimilation cycles performed with the Goddard Laboratory for Atmospheric Sciences (GLAS) general

circulation model (Somerville *et al.*, 1974; Stone *et al.*, 1977). The same model was also used in the forecasting experiments.

The model is a global, primitive-equation (PE) model, discretized in finite-difference form. Horizontal coordinates are longitude and latitude, the grid spacing being 4° latitude × 5° longitude. The vertical coordinate $\sigma = (p - p_s)/(p_s - p_T)$ is modified pressure, where p is pressure, p_s surface pressure, and $p_T = 10$ mb. The model has nine uniformly spaced levels in the vertical direction.

The horizontal space differencing is performed on a staggered grid, with quasi-conservation of mass, kinetic energy and enstrophy. Time differencing is performed following the space-centered Euler-backward (Matsuno) scheme (Arakawa and Lamb, 1977).

The model differs from that described in Somerville *et al.* (1974) and in Stone *et al.* (1977) by the introduction of a split grid, which at high latitudes doubles the longitudinal mesh size from 5° to 10° at 66°, and from 10° to 20° at 78° (M. Halem and G. Russell, personal communication). This modification is discussed in Herman and Johnson (1978). Its main effect is to allow an increase in the model time step from the previously reported 5 min to the current 10 min.

No explicit horizontal diffusion is added. Due to a number of features of the difference scheme, such as the averaging required by the staggering of variables and the additional smoothing at high latitudes, the scheme is in fact strongly dissipative. As a result, spurious high-frequency oscillations generated at the beginning of a forecast decay very rapidly, and reach a level comparable to that observed in the atmosphere after at most 6 h of simulated time. Hence, no special time-filtering or initialization procedure was used in either the assimilation experiments or in the associated forecast experiments.

a. The assimilation methods

The methods presented herein were designed to deal with temperature sounding data from polar-orbiting satellites. They can be extended, however, to other asynoptic meteorological data, such as cloud-track wind data from geostationary satellites, and aircraft, balloon and surface ship data. An application to Seasat-A marine wind data was given by Cane *et al.* (1978). The description will be restricted here to the sounding temperature application.

The vertical temperature profiles obtained from satellite-based radiance measurements are grouped by 10 min time intervals; the intervals are centered in time around a forecast model time step. A plot of a typical group of temperature data obtained in a 10 min interval is shown in Fig. 1. The choice of time interval corresponds to the length of the model time step. In what follows we shall refer to satellite-derived temperatures as *observed* temperatures.

All our methods carry out a correction of forecast temperatures at model grid points situated in the neighborhood of a group or "patch" of observed temperatures. This correction is applied to the forecast value at the model time step closest to observation time. It is based on interpolation of the observed-minus-forecast temperature field. The difference between methods consists in the way in which the interpolation coefficients are determined. In the present implementation of the methods, only information at the same mandatory pressure level is used, i.e., the interpolation is two-dimensional.

The general interpolation procedure is as follows. Let k be an observation point (on a fixed mandatory pressure level), T_k^o the observed temperature, and T_k^f the (interpolated) model temperature at the observation point k. We use a single subscript to indicate location, thus: $\mathbf{k} = (i,j)$, where i stands for discretized longitude and j for discretized latitude.

Let $\gamma_k = T_k^o - T_k^f$ be the difference between the observed and the forecast temperatures at k. We wish to compute corrections δ_l to forecast values at grid points $\mathbf{l} = (m,n)$ near the observation points k, where k ranges over a group or "patch" of observations such as the one shown in Fig. 1. In the figure, grid points closest to observation points k are marked by T's; additional grid points l which are affected by the observations after corrections are made appear as +'s. After applying the corrections calculated by interpolation, the temperature at + points as well as at T points will be different from the forecast value.

The corrections $\delta_l = T_l^c - T_l^f$, with T_l^c the corrected temperature at grid point l, are computed by the linear formula

$$\delta_l = \sum_k \alpha_l^k \gamma_k, \tag{1}$$

where here and in the following we drop the vector notation for k and l, the multi-index character of k and l being tacitly understood. The coefficients α_l^k in (1) are the coefficients of interpolation, or the "weights" which observations made at points k have in correcting the forecast value at point l. The difference between the assimilation methods stems from the different techniques used in determining the weights α_l^k.

1) The direct insertion method (DIM)

In this method α_l^k is different from zero only when the spherical distance s_{kl} between observation point k and correction point l is less than one model mesh width in either coordinate direction (cf. Bengtsson,

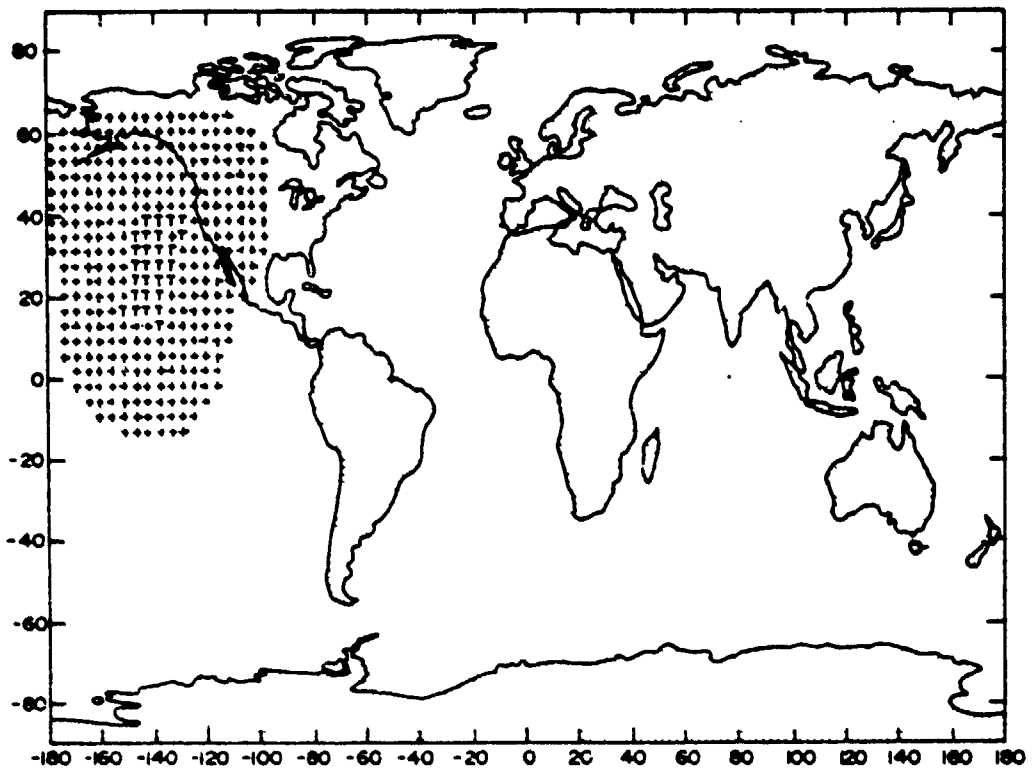


FIG. 1. The distribution of temperature sounding data available over a 10 mun time interval, and of the model grid points at which our statistical assimilation method will modify forecast temperatures using these data. The T's stand for grid points in the immediate neighborhood of satellite retrievals, the +'s for the other points affected by the retrievals in SAM. The total domain of influence of the retrievals, consisting of T points and + points, has a radius of 2000–3000 km.

1975, p. 24). The weights are calculated here as follows: within a model mesh rectangle given by its corners $l_0 = (i,j)$, $l_1 = (i+1,j)$, $l_2 = (i+1,j+1)$, $l_3 = (i,j+1)$, an observation point k defines four smaller rectangles, of areas A_0, A_1, A_2, A_3; the rectangle of area A_0 is adjacent to corner l_0, A_1 to l_1, A_2 to l_2 and A_3 to l_3. The areas A_0, A_1, A_2 and A_3 add up to the area A of the 4° latitude × 5° longitude mesh rectangle. The weight $\alpha_{l_0}^k$ is now simply set equal to

$$\alpha_{l_0}^k = \frac{1}{N_0}\frac{A_2}{A}. \tag{2}$$

Here N_0 is the total number of observations within the four mesh rectangles adjacent to the point l_0. Thus in particular $0 \leq \sum_k \alpha_{l_0}^k \leq 1$. The formula for the rectangle corner l_1 is $\alpha_{l_0}^k = (1/N_1)(A_3/A)$ etc. Further details on DIM appear in Ghil and Dilling (1978, pp. 3.4–3.10).

2) THE SUCCESSIVE CORRECTION METHOD (SCM)

Here the coefficients α_l^k are computed by an iterative procedure. The interpolation is carried out in a number of M successive scans. At each scan p, the interpolation extends to all observations within a radius R_p of the grid point l; R_p is the same for all grid points. The coefficients $\alpha_{l,p}^k$ at scan p are given by

$$\alpha_{l,p}^k = \begin{cases} \dfrac{1}{N_p}\dfrac{R_p^2 - s_{kl}^2}{R_p^2 + s_{kl}^2}, & \text{if } s_{kl} < R_p \\ 0, & \text{if } s_{kl} \geq R_p. \end{cases} \tag{3}$$

Here s_{kl} is the spherical distance between the points k and l, and N_p is the total number of observations within the circle with radius R_p centered at l.

We used $M = 4$ and $R_1 = 700$ km, $R_2 = 600$ km, $R_3 = 500$ km and $R_4 = 400$ km. As a result, for a given observation patch, the patch of *corrected* grid values will extend further out than in DIM. Moreover, due to the dependence of the weights α_l^k on distance, the corrected values will blend smoothly into the forecast field.

This method was implemented at NMC for the purpose of operational synoptic objective analysis by Cressman (1959). Further details on the asynoptic formulation used in this study appear in Ghil and Dilling (1978, pp. 3.23–3.32).

3) THE STATISTICAL ASSIMILATION METHOD (SAM)

In this method, the coefficients, or weights, α_i^k are determined by making explicit use of the error structure of the fields which we wish to interpolate. The knowledge of this error structure is incorporated into the correlation function $\phi(s)$. This function models the correlation between observed-minus-forecast temperature values at different locations.

(i) *The correlation function*. We have assumed that the error structure is *isotropic*, i.e., that ϕ depends on the distance s_{kl} between points k and l, but not on the direction of the position vector $\mathbf{r}_{kl}$. It is known that this assumption is not rigorously correct (Thiebaux, 1977) but it was convenient to make it at the outset in developing the current formulation of SAM [see Eq. (8) below and references following it].

The function $\phi(s)$ was computed first for discrete values of s, $s_1 = 100$ km, $s_2 = 300$ km, . . . , $s_p = (2p - 1) \times 100$ km, . . . , directly from the difference between satellite data and model forecast temperatures available during the DST-6 period. The continuous function $\phi(s)$ was then obtained by fitting an analytic function $\phi = \phi(s; s_0,c)$, depending on the parameters s_0 and c, to the values $\phi_p = \phi(s_p)$, by a least-squares fit. In other words, we obtained the values of s_0 and of c for which $\phi(s; s_0,c)$ satisfied

$$\sum_p [\phi(s_p; s_0,c) - \phi_p]^2 = \text{minimum}. \qquad (4)$$

The form of ϕ used in the experiments reported on here was exponential, i.e.,

$$\phi(s; s_0,c) = (1 - c)e^{-s/s_0} + c. \qquad (5)$$

Least-squares fits using a normal or Gaussian function with $\exp(-s^2/s_0^2)$ instead of $\exp(-s/s_0)$ gave residuals typically larger by an order of magnitude than those attained with (5). The distribution of ϕ_p did not seem to warrant the use of a series of Bessel functions J_0 (Rutherford, 1972) or of a damped cosine function (Schlatter *et al.*, 1977). Our experiments lead us to believe that results would not be particularly sensitive to the exact form of $\phi(s)$, or to the choice of parameters values.

The function $\phi(s; s_0,c)$ obtained in Eqs. (4) and (5) was not actually used in this form to calculate the weights α_i^k. It was modified in two ways: 1) to account for the known bias of remote-sounding temperatures; and 2) to account for the strong inhomogeneity of the error structure.

A positive value of c in (5) can be interpreted as being the result of a systematic bias in the differences $\gamma_k = T_k^0 - T_k^f$. Indeed, let the mean $\bar{\gamma}_k$ of the random variable γ_k be $\bar{\gamma}_k = \gamma \neq 0$. It is reasonable to assume that deviations from the mean of the differences γ_k are uncorrelated over large distances, i.e., that

$$\lim_{s_{kk'} \to \infty} \overline{(\gamma_k - \gamma)(\gamma_{k'} - \gamma)} = 0 \qquad (6a)$$

or that

$$\overline{\gamma_k \gamma_{k'}} \to \gamma^2 > 0 \qquad (6b)$$

as $s_{kk'} \to \infty$. This argument suggests that $c \approx \gamma^2$.

For most of the data sets used in our experiments (see Table 1 below) the values of c given by the best two-parameter least-squares fit were positive and small. A negative value of c was obtained for one of the sets of data considered here. The meaning of a negative c in this context is not entirely clear; indeed, $\gamma < 0$ would still give a positive $c = \gamma^2$. However, it is well known that negative values of $\phi(s)$ at large distances obtain for certain meteorological fields; they are associated with the statistical sampling of very long waves in the atmosphere (e.g., Gandin, 1963, Figs. 10 and 14; Phillips, 1976, Figs. 5 and 7; Schlatter *et al.*, 1977, and references therein). A more careful study of the effect of long waves on the statistical structure of meteorological fields remains to be done (e.g., Thiebaux, 1977 and references therein). At this stage a function $\phi(s; s_0,c)$ with $c < 0$ was used only for reasons of consistency.

In any case, these considerations show that an attempt to remove the observational bias requires modifying $\phi(s; s_0,c)$ so that $c = 0$. Hence the value of s_0 obtained in Eqs. (4) and (5) was kept, but c set to zero, i.e.,

$$\phi(s; s_0) = \phi(s; s_0,0) = e^{-s/s_0} \qquad (5')$$

was used instead of $\phi(s; s_0,c)$.

The spatial inhomogeneity of the error structure became apparent when computing stratified correlation values $\phi_p^n = \phi^{(n)}(s_p)$. Here n denotes a 10° wide latitude band within which pairs of differences were chosen, with n going from 1 to 18. The different curves $\phi(s; s_0^{(n)}) = \phi^{(n)}(s)$ showed a remarkably strong and systematic effect of latitude on the correlations (Fig. 2): the curves at higher latitudes had much narrower peaks, i.e., smaller $s_0^{(n)}$ than those at lower latitudes. This is in agreement with the well-known influence of latitude on the Rossby radius of deformation and on the mechanism of geostrophic adjustment. Indeed, the value of the radius of deformation affects the characteristic scale of both the forecast and the observed temperature fields and hence the spread of $\phi^{(n)}$. No sizeable or systematic dependence on height of the correlations was observed when stratifying computations by pressure levels.

Rather than base the correlation curve $\phi^{(n)}(s)$ used within each latitude band on the relatively small amount of data available within that band, we decided to model the effect of latitude in the following manner: A single correlation function $\phi = \phi(s; s_0,c)$ was computed as indicated before from

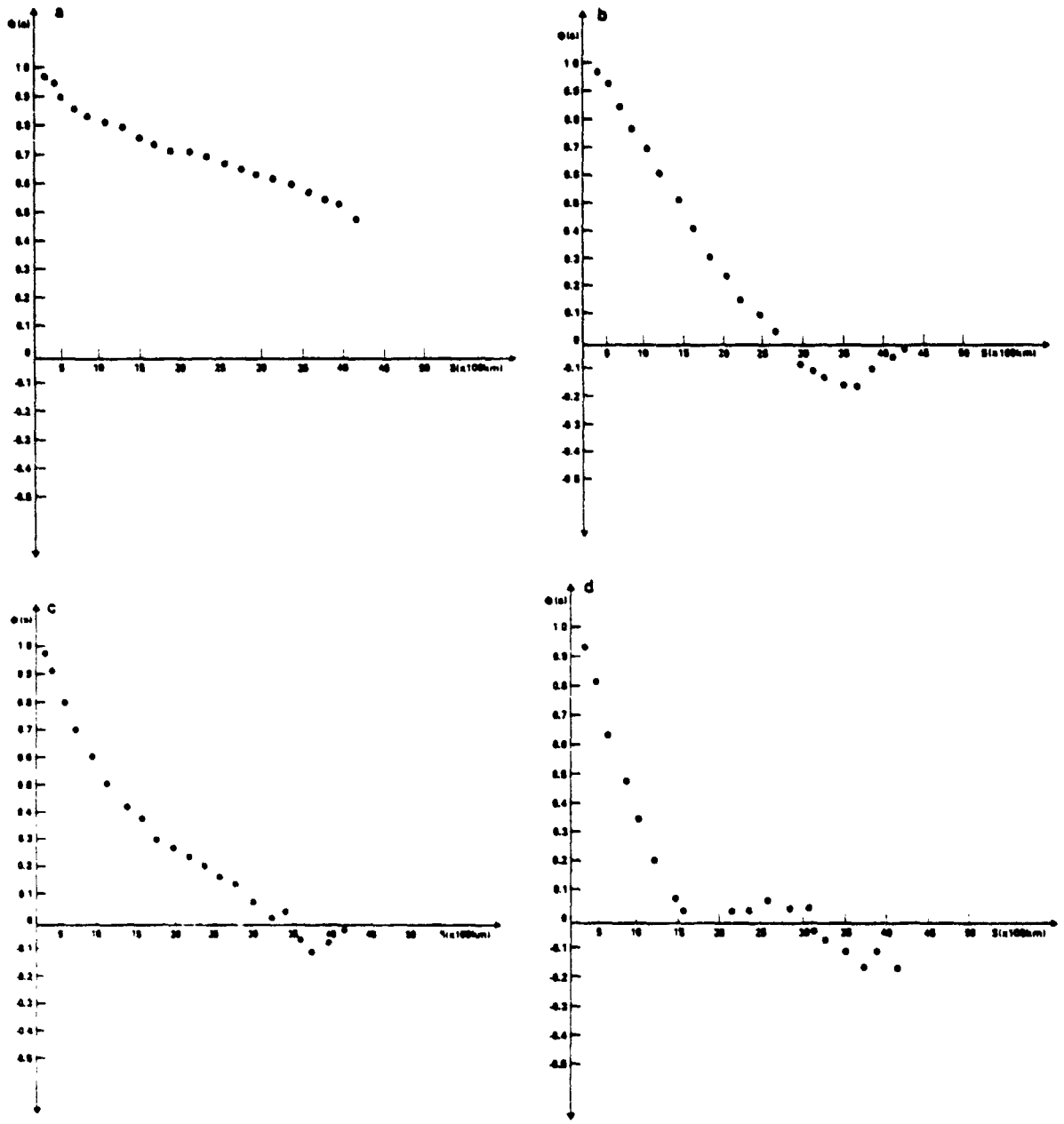


FIG. 2. The correlation function $\phi(s)$ for different latitude bands, using data at all the mandatory pressure levels. Clearly, the function has a narrower peak for higher latitudes. (a) 10°S–equator, (b) 40°S–30°S, (c) 40°N–50°N, (d) 70°N–80°N.

Eqs. (4) and (5), and modified by setting $c = 0$ to account for the bias [cf. Eq. (5')]. The function $\phi(s; s_0) = \phi(s; s_0, 0)$ was then modified further in order to allow for the latitude effect, so that finally the function actually used for a given observation patch was

$$\phi(s) = \phi(s; s_0 \cos\bar{\theta}), \tag{7}$$

where $\bar{\theta}$ is the mean latitude angle of the observation patch. The use of such a latitude-scaled correlation function has the added advantage that polar regions, over which the earth-orbiting satellites pass frequently, are not completely "swamped" by the sounding data which have lower accuracy there; such "swamping" could cancel the stabilizing influence of model forecast values and affect adversely the assimilation process.

(ii) *The linear regression equations*. Having obtained $\phi(s)$ in Eqs. (4)–(7), the weights α_i^k in (1) are de-

fined for this method as the solution of the system of linear equations

$$\mathbf{A}_l\boldsymbol{\alpha}_l = \boldsymbol{\beta}_l; \tag{8}$$

here $\boldsymbol{\alpha}_l = (\alpha_l^1, \ldots, \alpha_l^N)^T$, N is the number of observation points in the patch and $(\ )^T$ denotes the transpose of a vector or of a matrix. System (8) is the normal system which arises in all statistical applications using linear regression. Derivations of (8) can be found in the meteorological literature for instance in Bengtsson and Gustavsson (1971, 1972), Gandin (1963) and Rutherford (1972, 1973). It is based on minimizing the expected value of interpolation error squared with respect to the values of the weights α_l^k.

The statistical information accumulated on observations and on forecasts is incorporated into the entries $a_{kk'}$ of the matrix A and into the components β_l^k of the right-hand side vector $\boldsymbol{\beta}_l$. These are given by

$$a_{kk'} = \phi(s_{kk'}), \quad \beta_l^k = \phi(s_{kl}). \tag{9}$$

Here $s_{kk'}$ is the spherical distance between the two observation points k and k', and s_{kl} is the distance between the observation point k and the grid point l, at which we wish to make the correction. Notice that A is determined exclusively by the geometry of the observation patch, and we can drop the subscript l, while $\boldsymbol{\beta}_l$ depends on both observation points and the correction point we consider.

The assumption of an isotropic correlation function $\phi = \phi(s)$ leads to considerable simplification of the matrix A and to a much more reliable estimation of its entries $a_{kk'}$ from the available data. Hence it is quite customary to make this assumption (see references above and discussion of ill conditioning further below).

(iii) *Matrix inversion method.* The weights $\boldsymbol{\alpha}_l$ depend on the correction point l, because $\boldsymbol{\beta}_l$ in Eq. (8) does. This would seem to imply that for every correction point l we need to solve (8) for the weights $\boldsymbol{\alpha}_l$; such an approach would be computationally rather expensive. The number N of observation points in a patch is typically 50, so that a few thousand algebraic operations might be involved in making the correction at just one point. Fortunately, it is rather easy to circumvent this difficulty.

Let $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_N)^T$ stand for the vector of differences γ_k. Combining Eqs. (1) and (8), the correction δ_l is given by

$$\delta_l = \boldsymbol{\alpha}_l^T\boldsymbol{\gamma} = (\mathbf{A}^{-1}\boldsymbol{\beta}_l)^T\boldsymbol{\gamma}. \tag{10}$$

Here $\mathbf{x}^T\mathbf{y}$ is the inner or scalar product of the column vectors x and y, and $\mathbf{A}^{-1}$ is the inverse of the matrix A. Since A is symmetric, we can write

$$\delta_l = \boldsymbol{\beta}_l^T\mathbf{A}^{-1}\boldsymbol{\gamma} = \boldsymbol{\beta}_l^T\boldsymbol{\eta}; \tag{11}$$

here we used the rules for the transpose of a product and of an inverse, and introduced the vector $\boldsymbol{\eta} = \mathbf{A}^{-1}\boldsymbol{\gamma}$.

Eq. (11) makes it evident that we only need to compute the solution of system

$$\mathbf{A}\boldsymbol{\eta} = \boldsymbol{\gamma} \tag{12}$$

for an observation patch once; all the corrections δ_l are then computed as inner products $\boldsymbol{\beta}_l^T\boldsymbol{\eta}$. Thus, the total number of operations per patch is reduced to a few thousand, rather than a hundred times as much.

It is worthwhile noticing at this point that the device described in Eqs. (10) to (12) is useful in reducing the computational effort for any statistical assimilation method, whether time-continuous or intermittent. Indeed, consider in general N observation points k, and M points l at which corrections have to be made, based on the N observations; usually $M \geq N$. Then the device reduces the number of operations to *one* linear system solution (12), plus M inner products (11). On the other hand, *any* method of matrix inversion is equivalent to N system solutions, e.g., $O(N^3)$ multiplications for a full matrix. The number of inner products (10) is still the same as in (11). The net gain is thus considerable and increases as M and N increase.

In the current version of the method, a general-purpose linear equation solver is used to solve the system (12); the solver is based on Gaussian elimination with partial pivoting. In the future we intend to exploit the approximately banded structure of A in order to reduce the number of operations even further.

A well-known difficulty in solving system (8) or system (12) arises from the fact that such systems appearing in statistical computations are typically ill conditioned. In other words, as the number N of observations $\gamma_1, \ldots, \gamma_N$ increases, the "weighted corrections" $\boldsymbol{\eta}$ become more and more sensitive to errors in the data, i.e., to errors in A and in $\boldsymbol{\gamma}$.

Because of the ill-conditioning of this system, as well as in order to presmooth observational errors, the temperature sounding data were first averaged to the forecast model grid points. This averaging had the additional advantage of further reducing computation time. In the future we plan to use for this purpose a more elaborate form of smoothing, as described in Ghil and Dilling (1978, pp. 3.18–3.23), instead of straightforward averaging.

(iv) *Treatment of the sounding temperature bias.* We have already referred to the problem of systematic differences between the temperature fields obtained from satellite observations and those obtained from the analysis of conventional data; these differences appear whether climatology or model forecasts are used as a background field ("first guess"). This mean difference, or bias, should be accounted for in trying to extract the maximum

amount of useful information from satellite data. We have attempted to do so in two different ways.

First, the correlation function $\phi(s)$ was modified as described in Eq. (5'). Second, there was an attempt to take into account the effect of the bias on the temperature corrections themselves. One approach in this attempt consisted in computing the average difference between observations and interpolated forecast values over a satellite-data patch

$$\tilde{\gamma}(\tau) = \sum_{\tau \le t < \tau + \Delta t} \gamma_k(t)/N_\tau. \qquad (13a)$$

The summation here extends over a time interval $\Delta t = 10$ min and N_τ is the number of observations within this time interval. The corrections δ_l are then made using the modified differences

$$\gamma_k' = \gamma_k - p\tilde{\gamma}, \quad 0 \le p \le 1. \qquad (13b)$$

Thus $p = 0$ means no modification of the basic procedure described above, while $p = 1$ means that $\tilde{\gamma}' = 0$. In practice only $p = 0$, $p = 1$ were used, and no experiments with intermediate values were carried out. An experiment with $p = 1$ is marked as "bias removed at observation point" in Table 1.

Another approach in our attempt to remove the effect of the satellite-temperature bias was to "remove the bias at grid points". In this approach, the actual correction made at grid point l is

$$\delta_l' = \delta_l - \alpha^{(l)}\bar{\delta}, \qquad (14a)$$

with $\bar{\delta}$ defined as

$$\bar{\delta} = \frac{\sum_{l'} \delta_{l'}}{\sum_{l'} \alpha^{(l')}}; \qquad (14b)$$

here $\alpha^{(l')}$ is the sum of the weights at correction point l',

$$\alpha^{(l')} = \sum_{k=1}^{v_{l'}} \alpha_k^{l'}, \qquad (14c)$$

while the weights $\alpha_k^{l'}$ are given by (8). Notice that the modified corrections δ_l' sum to zero, i.e.,

$$\sum_l \delta_l' = 0, \qquad (14d)$$

but the modifications are not uniform over the whole domain of grid points influenced by a patch of observations. They are proportional to $\alpha^{(l)}$ and thus tend to be large toward the middle of the patch and small away from the patch.

This completes the description of the time-continuous SAM, as implemented in the experiments described in Section 3. It has already been presented in Ghil *et al.* (1977b); further details can be found in Ghil and Dilling (1978, p. 3.89 ff.).

(v) *Geostrophic wind corrections*. After the temperature corrections were carried out by either interpolation method (DIM, SCM or SAM), a local geostrophic wind correction was also performed. Such a correction was considered first by Hayden (1973); it is discussed by Bengtsson (1975, pp. 29–30) and by Kistler and McPherson (1975). In the model's σ coordinates it is given by

$$\mathbf{V}_l^c = \mathbf{V}_l^f + f^{-1}\hat{\mathbf{k}} \times \nabla_l(\phi^c - \phi^f), \qquad (15a)$$

where $\mathbf{V}$ is the horizontal velocity vector, f the Coriolis parameter, $\hat{\mathbf{k}}$ the vertical unit vector, ∇_l the gradient operator parallel to σ surfaces at point l, and ϕ the geopotential. Superscripts $(\ )^c$ and $(\ )^f$ stand for corrected and forecast values, respectively.

In the model, both ϕ^c and ϕ^f are computed from the hydrostatic equation

$$\frac{\partial \phi}{\partial \sigma} = -\pi/\rho, \qquad (15b)$$

where ρ is density and $\pi = p_s - p_T$, with p_s the local surface pressure and $p_T = 10$ mb. Corrected values of ρ, i.e., ρ^c, are computed from the equation of state, $\rho = p/RT$, using corrected temperature values T^c and model forecast pressures p^f.

No observations of p_s are available in general near the sounding-temperature "patches". Hence no corrections were made to the surface-pressure gradient term $(\sigma/\rho)\nabla\pi$ in evaluating our geostrophic wind corrections.

b. The evaluation methods

The effect of assimilating satellite data on weather forecasting was evaluated using the following three criteria: 1) differences between initial states produced by analyses benefiting from the use of satellite data, referred to as SAT initial conditions, and those produced without such data, *viz.*, NOSAT initial conditions; 2) differences in the numerical predictions made from these initial conditions, referred to as the SAT and NOSAT prognostic fields, respectively; and 3) differences in the actual weather forecasts which a local forecaster would produce based on acceptance of the information given by the SAT or NOSAT numerical prognostic fields. Differences in the prognostic fields were measured objectively by the standard numerical criteria of S_1 skill scores and rms errors, as well as by subjective comparisons based on synoptic criteria.

The qualitative and semi-quantitative comparisons involved in evaluating initial state differences, as well as the synoptic case studies made of prognostic fields, need no further elaboration at this point. Their use in Section 4 will be self-explanatory. We shall only describe here the analysis of S_1 and rms results for the numerical forecasts, and the automated forecasting method (AFM) used to produce local weather forecasts from the large-scale prognostic fields.

1) Numerical evaluation of forecast differences

The numerical measures of forecast accuracy used were S_1 skill scores and root-mean-square (rms) differences. The meteorological fields to which they were applied are the sea level pressure p_s and the height ϕ of the 500 mb geopotential surface. First the differences between values of the field produced by a model forecast and the field values of the NMC objective analysis at the same synoptic time were computed. These differences were computed for the synoptic time 48 and 72 h after initial time.

The next step was to compute the rms value of the difference field. The actual impact measure consisted in subtracting the rms difference for the experiment forecast from the rms difference for the NOSAT forecast. A positive value of this difference then means beneficial impact of, or improvement due to, satellite data for the given experiment and quantity, while a negative value means negative impact. The computations for skill score impacts were done in analogous fashion. The verification regions over which the impact was measured were North America (30–70°N, 75–130°W) and Europe (30–86°N, 10°W–40°E).

The computed values of S_1 impacts from the different experiments are given in Table 2, for the 500 mb height field ϕ verified at 48 h over North America. Similar tables were computed for S_1 skill scores and rms errors of the ϕ and p_s fields verified at 48 and 72 h over North America, as well as over Europe, i.e., 16 tables in all; they appear in Ghil and Dilling (1978, pp. 3.142–3.157) and will not be reproduced here in order to save space.

Considering the results from each experiment, one notices great variations in the daily values of forecast skill scores for all quantities, i.e., for both p_s and ϕ over both verification regions and at both verification times; negative as well as positive impacts occur, and they are of varying magnitude. These variations can be attributed to a number of factors; the quantity and quality of the satellite data during the 48 h preceding the given day figures prominently among these factors, along with the synoptic situation at the initial forecast time itself. Other factors can be simply assumed to generate random differences in the impacts.

Proceeding from Table 2, we shall now outline how the overview of results in Table 3 was arrived at. This is more a procedure for the concise presentation of a large quantity of results than a real statistical analysis model. In formulating a statistical model, one should separate first deterministic from random effects. The model should then account for the existing linear dependencies among the random effects. Such a model is currently under investigation.

For instance, when considering the random component in the outcome of our experiments, it is reasonable to assume that the results of one forecast are only weakly correlated with those of another, their initial times being separated by 48 h. It is also reasonable to assume in general only weak correlations between the outcome of verifications over Europe and those over North America, the separation being sufficiently large in space and hence in synoptic situations. But the results for ϕ on the one hand and for p_s on the other are strongly correlated, and so are results for the same quantity verified at 48 h and 72 h. The same is true of rms and S_1 error measures.

The synopses of results for rms errors and for S_1 skill scores have been compiled separately. They show strong correlation, and thus support our simplified analysis model of the detailed results.

In Table 2, let the entries in the columns labeled 2/1 through 2/21 be denoted by x_{ij}, with i the row number, $0 \leq i \leq M, M = 16$, and j the column number, $1 \leq j \leq N$, $N = 11$. Thus the actual skill scores for the NOSAT experiment NO are x_{0j}, $1 \leq j \leq N$, while the skill score impacts for all the other experiments are x_{ij}, $1 \leq i \leq M$; the same holds for rms errors.

The results in the three columns with the heading "Actual value" represent an overview of skill scores themselves, i.e., of x_{0j} for the first row, $i = 0$, and of $y_{ij} = x_{0j} + x_{ij}$ for the other rows, $i \neq 0$. The column headed "Ave" contains the averages

$$\bar{x}_0 = \sum_{j=1}^{N} x_{0j}/N \quad \text{and} \quad \bar{y}_i = \sum_{j=1}^{N} y_{ij}/N, \quad i \neq 0,$$

respectively; the column headed SD contains the standard deviations δ_i of x_0 or y_i, defined by

$$\delta_i^2 = \sum_{j=1}^{N} (y_{ij} - \bar{y}_i)^2/N, \quad i \neq 0,$$

and similarly for δ_0. The column headed SE contains the so-called standard errors $\hat{\delta}_i$ in estimating the means $\bar{x}_0$ and $\bar{y}_i$, $i \neq 0$, respectively: it is computed as $\hat{\delta}_i = \delta_i/\sqrt{N}$.

The results in the three columns headed "Difference from NO" represent an overview of the impacts x_{ij}, $0 \leq i \leq M$. The entries in the first row are therefore clearly zero. The column headed "Ave" contains $\bar{x}_i = \sum_{j=1}^{N} x_{ij}/N$, column SD contains $\sigma_i, \sigma_i^2 = \sum_{j=1}^{N} (x_{ij} - \bar{x}_i)^2/N$, and finally SE contains $\hat{\sigma}_i = \sigma_i/\sqrt{N}$.

Because of the random influences on the detailed results we discuss, there were still great differences between the results in the last three columns of Table 2 and those of the similar tables for the other quantities, verification areas and verification times (Ghil and Dilling, 1978, pp. 3.142–3.157). The results in the different tables were further summarized as follows.

The average impact $\bar{x}_i$ of a given experiment was

computed from the values x_{ik} in the eight different tables with skill score results (Tables 39a–39h, *loc cit.*), $\tilde{x}_{ik} = (\hat{x}_{ik} - \hat{x}_{0k})/\hat{x}_{0k}$, $1 \leq k \leq 8$. We simply set $\bar{x}_i = \sum_{k=1}^{8} \tilde{x}_{ik}/8$, and will refer in the sequel to $\bar{x}_i$ for convenience as x_i. We shall use this average x_i as a measure of overall impact for experiment i, together with the corresponding rms impact.

Computing the statistical significance of the mean impact or estimating its variance is a delicate matter. Here the assumptions about the population out of which the table entries are considered to be a sample play a critical role.

We define the measure of statistical significance, or confidence in our results, in the following operational way. In each one of the eight tables with S_1 results, we compute $\tilde{\zeta}_{ik} = \tilde{x}_{ik}/\hat{\sigma}_{ik}$, $1 \leq i \leq M$, $1 \leq k \leq 8$. The total measure of significance of any given experiment i is then given by simple averaging, as for x_i, $\bar{\zeta}_i = \sum_{k=1}^{8} \tilde{\zeta}_{ik}/8$. The values $\bar{\zeta}_i$ are related to a standardized score or z score for the mean of the impacts, in terms of a probability distribution with zero mean and unit standard deviation. In this treatment we have neglected the uncertainty in estimating the standard deviation of the mean impact; this uncertainty would be taken into account in the careful application of a Student's t test to the confidence limits of the mean. At this point suffice it to say that for a number of degrees of freedom ν, or sample size N, $\nu = N - 1 = 10$, the t distribution is practically undistinguishable from a standardized normal distribution; the remark is valid *a fortiori* if we are willing to grant that verification results over Europe are independent of those over North America, and hence $\nu = N - 1 = 21$.

After these remarks, we shall limit ourselves to stating that we have very little confidence in the results of a given experiment i if $|\bar{\zeta}_i| \leq 0.5$, reasonable confidence if $0.5 < |\bar{\zeta}_i| \leq 2$, and very high confidence if $2 < |\bar{\zeta}_i|$. The values 0.5, 1 and 2 of $z = \bar{\zeta}_i$ correspond to the following values of the cumulative distribution function $F(z)$ of a normal distribution with zero mean and with standard deviation unity:

$$F(0.5) = 0.69, \quad F(1) = 0.84, \quad F(2) = 0.98.$$

In the following, $\bar{\zeta}_i$ will be referred to as ζ_i, for notational convenience.

This completes the description of the procedure for summarizing the S_1 and rms results of our forecast experiments.

2) The automated forecasting method (AFM)

The practical usefulness of NWP depends to a large extent on its ability to produce or help in producing local weather forecasts, especially precipitation forecasts. Vertical motion and relative humidity play a crucial role in such local forecasts. Unfortunately, the large-scale fields of these quantities are not predicted with sufficient accuracy by current NWP models.

Quasi-geostrophic theory provides a way of computing the vertical motion field with satisfactory accuracy from a knowledge of the three-dimensional geopotential field, using the ω equation. This equation, together with the geopotential tendency equation, are the classical tools for the diagnostic analysis of synoptic-scale midlatitude disturbances (e.g., Holton, 1972, Chap. 7).

The basic idea of the AFM is to apply these tools to the diagnostic analysis of the numerical model's *forecast* fields. The geopotential field and the wind field are among the fields most accurately predicted by the NWP model; the latter also yields realistic estimates of actual vorticity. Hence one can expect more accurate estimates of vertical motion from this procedure than those provided directly by the model. Such a procedure also conforms to subjective forecasting practices which are widely used in local weather prediction.

The development of the Automated Forecasting Method is an attempt to quantify the usefulness of large-scale numerical forecasts in local weather forecasting. The AFM was designed to simulate the subjective interpretation of a set of prognostic charts under the restrictive condition that no modification of the prognostic flow patterns, as forecast by the numerical model, can take place. It has two main advantages for objective evaluation: 1) local precipitation forecasts can be generated and analyzed for a large number of locations in a very short time, and 2) the procedures that are applied to each set of prognostics will be uniform.

The major quantities forecast by the AFM are precipitation occurrence, type of precipitation and 24 h surface temperature change. We shall restrict ourselves here to a description of the AFM precipitation scheme, which was actually used in evaluating results from some of the DST-6 experiments carried out.

The first step in the method is to compute several forecast parameters needed for every location at which a local forecast is to be made. The forecast parameters used by the method are absolute vorticity advection (VA) at 500 mb, temperature advection (TA) at 850 mb, advection of the 1000–500 mb thickness (THA), moisture advection (MA) at 850 mb, dew-point depression (DD) at 850 mb, wind direction (WD) computed geostrophically at sea level, vertical wind shear (VWS) from sea level to 850 mb and from 850 to 700 mb, and 850 mb temperature (T850). This list includes most of the quantities used by operational forecasters to predict areas of large-scale precipitation (Houghton and Irvine, 1976). For instance, vorticity advection at 500 mb is a reasonable approximation to the differential vorticity advection term in the

ω equation, since vorticity advection at the surface is very small.

Local quantities, such as T850, are computed by linear interpolation from the corresponding grid point quantities. Advection of a quantity Q is defined as $(\mathbf{V}\cdot\nabla)Q$, where $\mathbf{V}$ is the horizontal velocity vector in pressure coordinates and ∇ the corresponding horizontal gradient operator.

The numerical calculations involved in computing $(\mathbf{V}\cdot\nabla)Q$, such as averaging and differencing, were all based on the four grid points adjacent to the desired location; since absolute vorticity η is not computed directly by the model, 12 points, rather than 4, were used in the calculations for $(\mathbf{V}\cdot\nabla)\eta$. The use of actual wind and actual vorticity in this term, rather than of their geostrophic values, was deemed more compatible with the PE model; it should lead to better estimates of vertical motion even within the framework of applying quasi-geostrophic relationships to the diagnosis of the model forecast results.

The next step is to determine whether there is sufficient moisture at the given location for precipitation to occur there. If this criterion is met, the vorticity advection field is tested. Positive vorticity advection (PVA) is an indicator of upward vertical motion, whereas negative vorticity advection (NVA) is an indicator of downward vertical motion. Generally, strong PVA in the presence of sufficient moisture will result in precipitation. The effects of thermal advection are coupled with those of vorticity advection in the ω equation and strong cold advection can negate the effect of PVA, whereas warm advection will enhance it. The manner in which the AFM combines these quantities is illustrated in the Appendix (Fig. A1).

Once the above precipitation calculation is completed, two additional tests for precipitation are performed. These include a test for cold frontal precipitation (Fig. A2), and a gross check for the potential for showers (Fig. A3). Finally, a test for the type of precipitation is made as illustrated in Fig. A4. A sample of the AFM's output is presented in Fig. A5.

The AFM cannot perform pattern recognition in the same manner as a human forecaster. It does take into account, however, most patterns in terms of the computed advection of certain quantities; this process is somewhat analogous to that of graphic differentiation of the relevant prognostic fields.

Prior to its application to the evaluation of satellite data impact, the AFM was tested from two standpoints: its accuracy when applied to a perfect prognosis, i.e., to an analysis, and its agreement with subjective forecasts.

Comparisons with observed precipitation events were made for the month of February 1976, making precipitation forecasts twice a day at 11 cities, i.e., $2 \times 29 \times 11 = 638$ forecasts in all, from the NMC analysis valid at the corresponding synoptic time. The statistics verified were 1) *relative frequency* of correctly forecasting whether or not precipitation would occur, 2) *post-agreement*, i.e., the relative frequency of cases in which precipitation actually occurred when forecast by the AFM, and 3) *prefigurance*, i.e., the relative frequency of cases in which an observed precipitation event was forecast by the AFM. Post-agreement can be defined as (area correct)/(area forecast) and prefigurance as (area correct)/(area observed) (Tracton and Stackpole; in Murphy and Williamson, 1976, p. 736). Clearly, a precipitation event is labeled "correct" in this context if it is both forecast and observed. For the 638 cases tested, the AFM had a 76% verification in prefigurance, as well as in post-agreement, and a 93% verification in forecasting both positive and negative precipitation events; the latter score is obviously higher since absence of precipitation is more often forecast correctly.

The agreement with subjective precipitation forecasts was tested for part of the same period, using ten 72 h numerical forecasts and three forecasters working independently. The numerical prognostic fields from the 10 forecasts were used to make precipitation forecasts at five cities every 12 h, i.e., $5 \times 10 \times 6 = 300$ forecasts. The agreement between the AFM and the subjective forecasts was 94% on the average, for all three forecasters; it was comparable to the agreement between one forecaster and another. We concluded from these tests that the AFM was quite comparable in performance to the average skilled weather forecaster, and decided to use it in evaluating the numerical forecasts made in our DST-6 experiments.

Further details on the AFM can be found in Atlas and Sakal (1978) and in the Appendix.

3. Description of experiments

The purpose of our experiments was to study the effect of satellite data on the accuracy of initial states obtained with the aid of such data, and on the accuracy of forecasts starting from these initial states. Specifically, we studied the effect of the quantity and accuracy of the satellite data themselves, on the one hand, and of the assimilation methods used to extract the information from the data, on the other.

All the experiments consisted of a continuous assimilation run, extending over the entire period for which data were available, and of a number of forecasts started from selected initial states produced by the assimilation run. The experiments differed from each other by the sounding data which were assimilated, and by the method which was used to carry out the assimilation. The assimila-

tion methods used for the satellite data were time-continuous direct insertion (DIM), asynoptic successive correction (SCM) and time-continuous local statistical assimilation (SAM); in addition, an experiment with intermittent insertion at synoptic times only was also performed.

a. The data

The methods were applied to data from the DST-6 (January–March 1976) period. A complete assimilation cycle was carried out for all experiments from 0000 GMT 29 January to 0300 GMT 21 February 1976; for some of the experiments, the assimilation was continued to the end of the DST period, i.e., to 0300 GMT 4 March 1976. Eleven forecasts were carried out for all experiments, starting at 0300 GMT on 1, 3, 5, . . . , 21 February; for the extended experiments, six additional forecasts were performed, starting on 23, 25, 27 and 29 February, and on 2 and 4 March. The reason for starting the forecasts on 0300 GMT and not at 0000 GMT was to achieve as close a parallel as possible with current operational practice at the National Meteorological Center (NMC) because NMC uses intermittent assimilation with a ±3 h "window" for satellite data. Thus, an operational NMC forecast started at 0000 GMT uses all asynoptic information up to 0300 GMT; so do our experimental runs started at 0300 GMT as we rely on time-continuous assimilation. Starting 3 h after the insertion of synoptic data also has the effect of letting the model smooth out the initialization shock occurring at synoptic time (see Section 2; also Ghil, 1975, and Ghil *et al.*, 1977a).

The temperature sounding data were supplied by the National Environmental Satellite Service (NESS). They had been obtained by operational retrieval methods (Smith and Woolf, 1976; Werbowetzki, 1975) from the radiance measurements performed during DST-6 by the VTPR instrument aboard the NOAA 4 satellite and by the HIRS and SCAMS instruments aboard Nimbus 6. Different retrieval methods have been under development at the Goddard Space Flight Center as well (Susskind and Edelmann, 1978), but their results will not be discussed here.

The correlation parameters s_0 and c of Section 2 [Eqs. (4) and (5)] were computed separately for data from the two satellites, and for the two parts of the DST-6 period we considered, 30 January–20 February and 21 February–4 March. Then their values were also computed for the data from both satellites and from the entire period. The value of s_0 for data obtained with the VTPR instrument is almost twice as large as that for the HIRS instrument (2940 km vs 1781 km, for data from the entire period). But values for data from either satellite differ only very slightly from one part of the period to the other (e.g., from 1853 km in the first part to 1686 km in the second part for Nimbus 6. The values of s_0 and c actually used in the experiments are given in Table 1.

The sounding data from the first and second part of the DST-6 period do differ, however, in other respects. This is due to a change in the observational properties of the VTPR instrument aboard NOAA 4 during the period (H. Bowman, private communication). Also some aspects of the operational processing of Nimbus 6 data by NESS were changed (C. M. Hayden, personal communication).

The conventional data and experimental data other than temperature soundings were provided by NMC, in accordance with the data formats described in Automation Division Staff (1973). The data categories used were surface reports from land and ocean stations, including fixed and moving ships; upper air reports from land and ocean stations, as for surface reports, and in addition from conventional and special aircraft, from the Tropical Wind Energy Conversion Reference Level Experiment (TWERLE), and cloud-track winds obtained by NESS from the Synchronous Meteorological Satellite (SMS). Data categories not used were those from reconaissance aircraft, manual monitoring bogus data produced by NMC based on cloud pictures, and satellite winds obtained from SMS by the Space Science and Engineering Center at the University of Wisconsin, using cloud-picture processing procedures different from those of NESS.

Surface data used were all those within 3 h of synoptic times (from 2100 to 0300 GMT and from 0900 to 1500 GMT). The upper air data used were also restricted to these two 6 h "windows", and furthermore to latitudes lower than 85° in both hemispheres. All these data were assimilated at the synoptic times by direct insertion (DIM).

b. The methods

A summary of the experiments is given in Table 1. The null experiment, in which only synoptic data and no sounding temperatures were inserted, is denoted by NOSAT. It was performed at the beginning of this study (Experiment N0). Because of difficulties in the maintenance of a data base and of a program having the size involved in these experiments, the null experiment was repeated at the end of the testing period reported on in Halem *et al.* (1978); this rerun of the NOSAT case is Experiment N0r.

Experiment CNc was set up as a control experiment to check the information content of satellite data; it should show whether the smoothing effect of SCM or SAM on the field values produced by assimilation runs has an impact comparable to that of systematic differences between observations and forecasts or not. In this experiment the temperature data used as "observations" were model

TABLE 1. Summary description of impact experiments. The use of data from the VPTR instrument on board the NOAA 4 satellite is denoted by VPTR, the use of data from the HIRS and SCAMS instruments on board the Nimbus 6 satellite by NIMB. The method by which the data have been assimilated is indicated by the acronyms DIM, SCM or SAM. The experiments in which only conventional synoptic data, and no satellite data at all were used appears as NOSAT. The first character of the code number is a capital letter and refers to the method. DIM = D, SCM = C and SAM = S, the second character refers to data source: VTPR = V, NIMB = N and VTPR + NIMB = 2 (for 2SAT); the third character, a lower case Latin or Greek letter, refers to slight variations in the method. Additional explanations concerning the methods are given in the text.

Experiment	Assimilation method	Data		Correlation parameters	
		VTPR	NIMB	s_0 (km)	c
NO	NOSAT	0	0		
NOr	NOSAT[a]	0	0		
DN	DIM		X		
CNc	SCM[b]		X		
CN	SCM		X		
CN/	SCM[c]		X		
C2	SCM	X	X		
C2t	SCM[d]	X	X		
C2i	SCM[e]	X	X		
SV	SAM	X		2293	−0.124
SNa	SAM[a]		X	1831	0.018
S2	SAM	X	X	1842	0.015
S2a	SAM[a]	X	X	1842	0.015
S2b	SAM[b]	X	X	1842	0.015
S2μ	SAM[c]	X	X	1842	0.015
S2ν	SAM[c]	X	X	1842	0.015

[a] Rerun or repeat of Experiment NO.
Control experiment with simulated Nimbus 6 data, based on NOSAT (NO) 12 h forecast.
No insertion of satellite data over data-dense land areas.
Intermittent insertion at synoptic times of all data.
Opt mized (tuned) scanning radii.
Bias removed at observation point.
Bias removed at grid point.
Weighting of forecast and observation values according to statistical confidence, the difference between μ and ν is explained in the text.

values produced from a NO (see Table 1) 12 h forecast, to which simulated observed-minus-forecast differences had been added. These simulated differences were computed at the true Nimbus 6 observation locations, by using a random-number generator function with statistical properties determined from actual data.

The question of the relative reliability of satellite observations versus that of forecast values was specifically addressed in Experiments S2μ and S2ν. Given the values of the temperature corrections δ_l as computed in Section 2 [Eqs. (8)–(12)], the corrected values T_l^* of temperature at grid point l can be expressed as

$$T_l^* = T_l^f - \mu\delta_l. \tag{16a}$$

In all other SAM experiments $\mu = 1$. In these two experiments,

$$\mu = \sigma_f^2/(\sigma_f^2 + \sigma_o^2), \tag{16b}$$

which corresponds to a weighting of forecast and observation values in accordance with a statistically measured confidence in these values. The variances σ_f and σ_o were computed separately for 12 latitude bands and for 11 mandatory pressure levels, using colocated radiosonde measurements for the "true" value of the temperature field. In regions where a sufficient number of values were lacking for a reliable estimate of variance to be made, certain default options were introduced. One such default option was to use σ_o computed for a larger region, in which sufficiently many observations were available, and which included the given region.

The difference between Experiments S2μ and S2ν is in the way in which the presence within one updating interval of temperature retrievals from both satellites, NOAA 4 and Nimbus 6, is accounted for. The variances σ_o are computed separately for either satellite. In Experiment S2μ, μ for a "mixed patch" was computed as a weighted average of the μ values for the two satellites. In Experiment S2ν, a weighted average of σ_o for the two satellites was computed in a "mixed patch" and μ computed from (16) with this averaged value of σ_o.

The time-continuous approach to 4D assimilation of temperature sounding data was used in all the experiments described heretofore. In order to verify the validity of this approach, Experiment C2i was carried out. In it, the sounding data, as well as all the other data, were inserted intermittently at synoptic times; a successive correction method was used for both conventional and sounding data. The sounding data were grouped in 12 h windows and weighted according to their spatial distance and time separation from radiosonde observations, receiving less weight when close to the latter and more when far from them. A detailed report describing variance data for Experiments S2μ and S2ν, and the weighting procedure for Experiment C2i is in preparation; the temporal weight factor used is similar to the one in Ghil and Dilling (1978, p. 3.7).

The SCM experiments C2i and C2t used the scanning radii and computation of weights described in Section 2, the weights in these two experiments can be considered close to optimal, due to adequate tuning. In the other SCM experiments, CN, CN/ and C2, an error in the programming for Eq. (3) led to weights which were found to be still reasonable, but far from optimal. The difference in results between C2 and C2t will be discussed in Section 4.

Our timing estimates for the different assimilation methods show that the computational cost of implementing more sophisticated assimilation methods is not exaggerated. For example, a 24 h forecast takes 40 min of CPU time; a 24 h NOSAT assimilation runs in 48 min; a 24 h time-continuous SCM assimilation runs in 59 min; and a 24 h SAM runs in 96 min [cf. Ghil and Dilling (1978), where details are provided]. These estimates all refer to the present version of the GLAS GCM running on an

Amdahl 470/V6 computer with a core memory of 2 megabytes. In fact, the computational cost of SAM can be considerably reduced by using more efficient numerical methods and programming techniques.

4. Discussion of results

The experiments described in Section 3 will now be discussed according to the criteria of Section 2b.

a. Initial state differences

In the present context, the utilization of satellite data only affects the forecasts as a result of differences between the initial states computed with (SAT) and without (NOSAT) the aid of these data. Different initial states lead then to different forecasts. Therefore, we start by considering the differences in initial states produced by the use of satellite data. Initial state differences are presented between Experiment S2a and the NOSAT Experiment NO. We recall that in Experiment S2a data from both the NOAA 4 and Nimbus 6 satellites were assimilated by SAM: S2a is taken as representative of all the SAM experiments using data from both satellites (see also Section 4b below).

Figs. 3a and 3b show the 500 mb geopotential height difference fields for the two assimilation cycles, one including all the sounding data from both satellites (called 2SAT) and the other withholding all satellite soundings (called NOSAT). The difference fields are given at the initial time of each one of the forecasts which were carried out for every experiment. A look at the fields indicates that large variations occur in the magnitude of initial state differences from one day to another.

The magnitudes of the differences vary as a result of variations in data coverage and quality, as well as of the natural variability of the atmosphere. It is clear from the difference plots that day after day, differences of the order of 30–120 m are produced in data-sparse regions of both the Pacific and Atlantic Oceans. Geopotential height differences of this magnitude correspond to 1.5–6°C temperature differences in the 1000–500 mb thickness. The variations are of both signs, although there seems to be a systematic tendency for warmer temperatures in the SAT cycle over most of this assimilation test.

Superimposing the satellite tracks (Fig. 4) on the difference charts, we notice that the largest differences occur mainly in the data-sparse areas along the tracks where sounding data have been inserted recently, i.e., in the 6 h preceding synoptic time. It was also observed that in cases where the initial state differences are smaller, the forecast impact in terms of S_1 skill scores and rms errors is generally also smaller. We intend to substantiate these qualitative observations by further quantitative studies.

We present in addition initial state differences between the 2SAT SCM Experiment C2t and NO, for comparison purposes (Fig. 5). It is clear that, day by day, the differences tend to be somewhat smaller than those between S2a and NO. This is probably due to the difference between the methods. The warm bias is still apparent and would thus seem to be due to the data themselves, rather than to the method.

From the nature of the initial state differences, one cannot say *a priori* whether the utilization of satellite data in a given case would produce beneficial impacts. It seems reasonable to assert at this point, however, that the magnitude of these differences, in the case of SAM and of correctly tuned SCM assimilation, is sufficiently large so that one would expect them to produce a certain number of significantly distinct forecasts.

b. Numerical evaluation of forecast differences

Numerical measures of forecast accuracy for our experiments are summarized in Table 3. Before discussing these results, it is important to remember that Experiment CNc is a control experiment, as described in Section 3, and that all the results have to be gauged against those of Experiment CNc. The mean percentual impacts for CNc were 0.97 in S_1 scores and 1.51 in rms errors; the statistical significance was marginal, i.e., close to 1 for both S_1 and rms results.

1) Impact of data

The influence of data quantity can be assessed by comparing the results of experiments using the same assimilation method. Such a series of experiments is SV, SNa and S2, in which SAM was used to assimilate data from the NOAA 4 satellite only (SV), from the Nimbus 6 satellite only (SNa), and combined data from both satellites (S2, S2a and S2b). The SAM experiments S2, S2a and S2b gave very similar results (see discussion further below) and these were averaged into results representative of all three, denoted by S2m.

The sequence of S_1 mean percentual impacts is $r_{SV} = 1.20$, $r_{SNa} = 3.11$, $r_{S2m} = 4.66$. These are roughly proportional to the quantity of satellite data used in each experiment, and $r_{SV} + r_{SNa} \approx r_{S2m}$. A similar statement holds for rms mean percentual impacts. The statistical significance for all these results was high, with the exception of SV, where it was marginal.

A pair of comparable SCM experiments are CN and C2, with Nimbus 6 data and data from both satellites, respectively. The results for this pair are not quite conclusive. The S_1 impacts are $r_{CN} = 1.83$

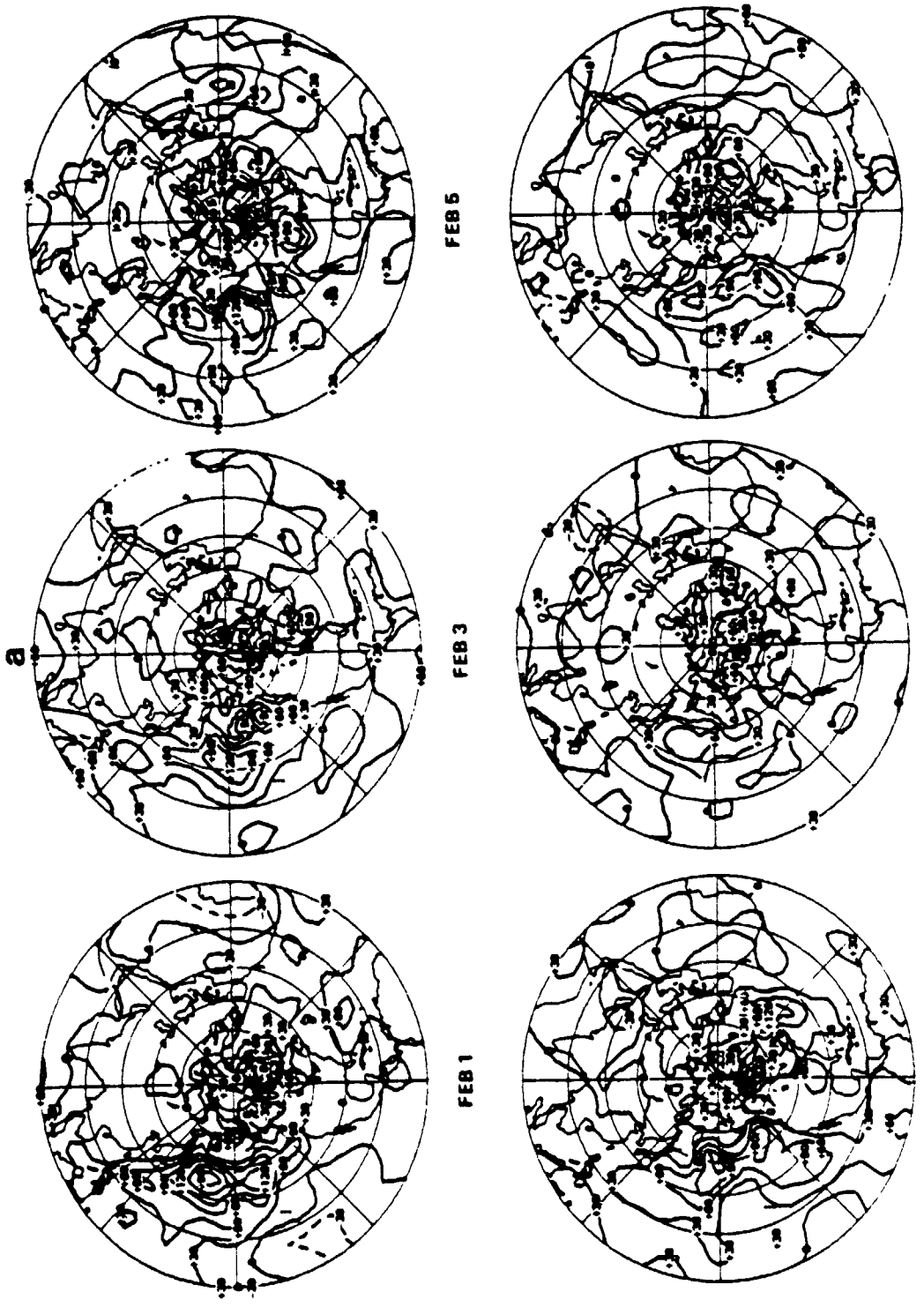


FIG. 3 Initial-state differences in 500 mb geopotential heights for the Northern Hemisphere between experiments S2a (2SAT/SAM) and NO (NOSAT), $\phi_{S2a} - \phi_{NO}$. Positive differences are indicated by full contour lines, negative differences by dashed lines. The differences are given at 0000 GMT. They show a definite warm bias of satellite sounding temperatures, especially over the oceans. (a) Differences from 1 through 11 February 1976, (b) differences from 13 through 21 February 1976.

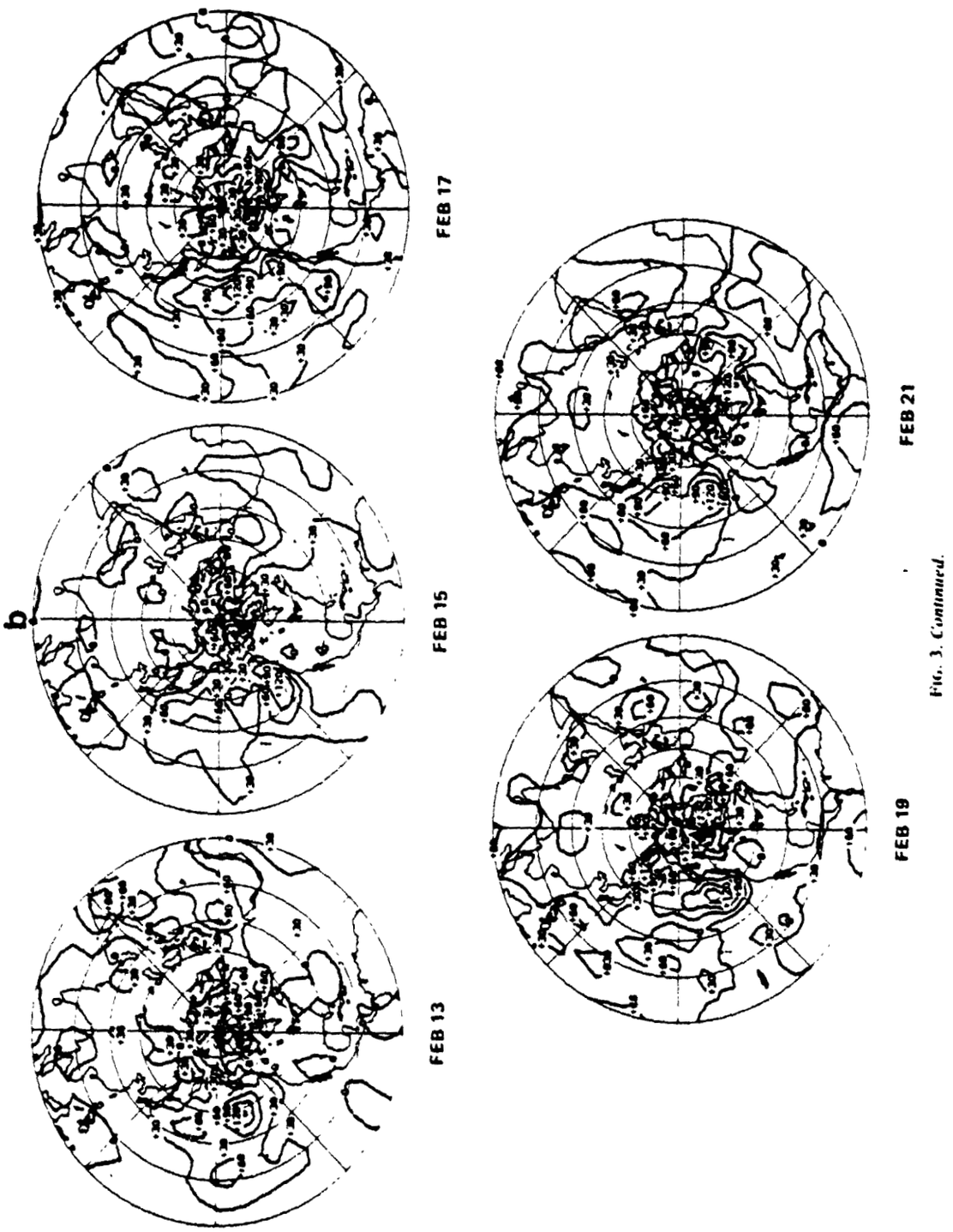


FIG. 3. *Continued.*

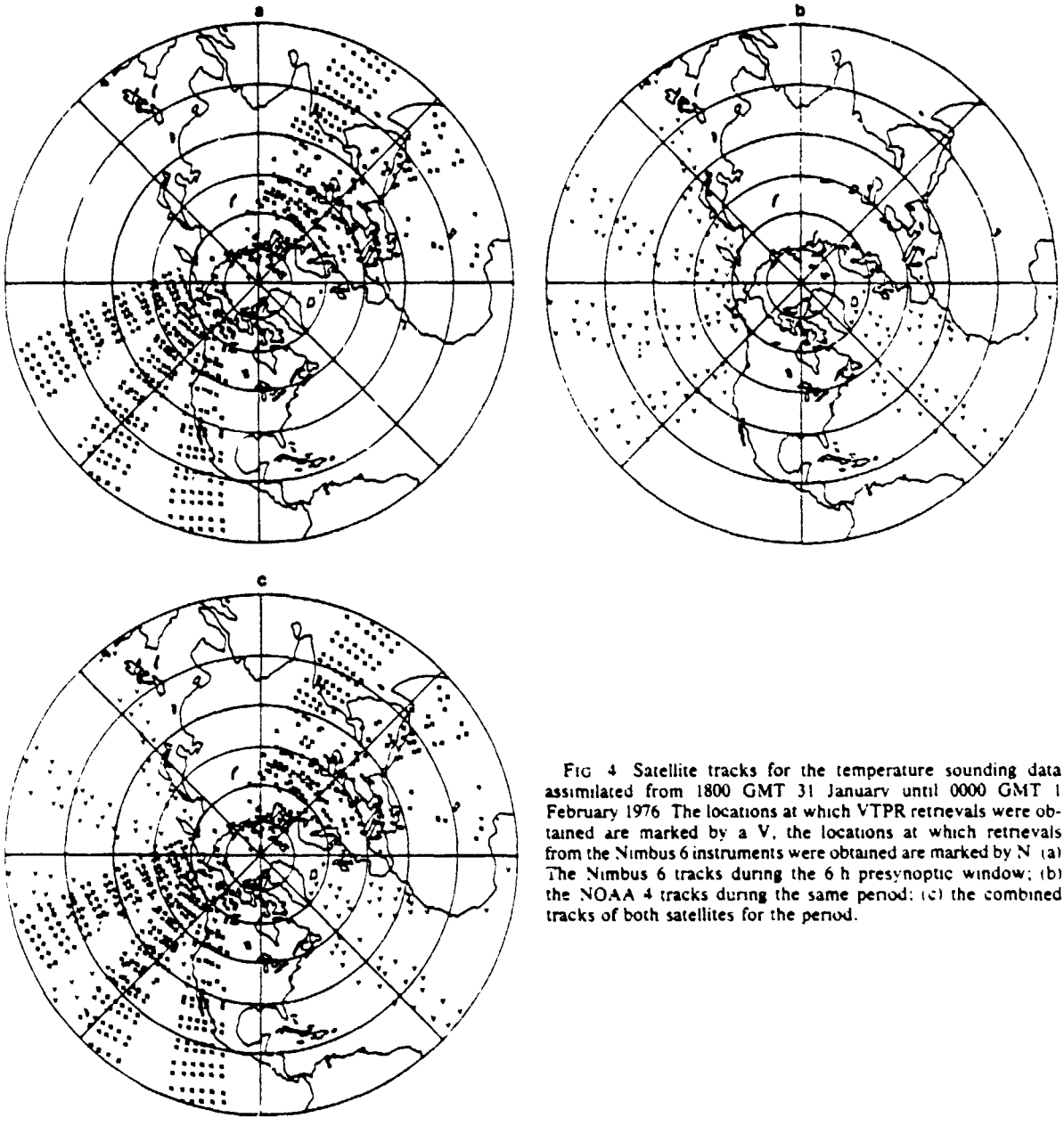


FIG. 4. Satellite tracks for the temperature sounding data assimilated from 1800 GMT 31 January until 0000 GMT 1 February 1976. The locations at which VTPR retrievals were obtained are marked by a V, the locations at which retrievals from the Nimbus 6 instruments were obtained are marked by N. (a) The Nimbus 6 tracks during the 6 h presynoptic window; (b) the NOAA 4 tracks during the same period; (c) the combined tracks of both satellites for the period.

and $x_{C2} = 2.75$, i.e., a 50% improvement in mean percentual impact. However, rms results are not entirely consistent with S_1 results.

In any case, we are led to conclude that independent of assimilation method, the quantity of satellite data used affects the numerical measures of impact on forecasting accuracy. The impact seems to increase with data quantity; this increase is even more pronounced when the satellite data are assimilated by a method which explicitly takes into account their error structure (SAM).

The results discussed above, and all the results in Table 3, refer to forecasts from the first part of the DST-6 period, 1–21 February 1976. Experiments NO, NOr, S2a, S2μ and S2ν were extended to the end of the DST period, 4 March 1976. The results of the extended SAM experiments, S2a, S2μ and S2ν are given in Table 4. The correlation parameters used for the second part were the same as for the first part (see Table 1).

In Table 4 a comparison is made between results for the first part of the period, the second part and the entire period. It is clear that the results for the second part are considerably worse than for the first part. We believe that this is due to the changes in data acquisition and processing procedures, and

not to the change in the statistical error structure, which was small (cf. Section 3).

The deterioration of results during the second part of DST-6 seems to stress the importance of data quality, as well as data quantity. A method, even though it takes into account explicitly error structure (S2a, S2μ and S2ν), cannot entirely compensate for presence of error in the data.

2) Impact of methods

The methods used for assimilation can be classified as intermittent, for experiment C2i, and time-continuous, for all the other experiments. Among the time-continuous experiments the distinctions are given by the way interpolation coefficients are computed (Section 2a); the main differences are between direct insertion (DIM), asymptotic successive corrections (SCM) and local statistical assimilation (SAM). Within each method, further differences arise and will be discussed.

Experiment C2i, based on intermittent assimilation, is in all other respects closely comparable to the time-continuous SCM experiment C2t: it uses data from both satellites and properly tuned weighting coefficients. For S_1 scores, the mean percentual impacts were $x_{C2i} = 2.79$ and $x_{C2t} = 5.01$. The corresponding numbers for rms errors are $x_{C2i} = 3.28$ and $x_{C2t} = 9.31$, an even larger increase of impact due to the time-continuous approach. Statistical significance of all these results was good, i.e., larger than 1.33 for C2i and larger than 2.10 for C2t.

The results of such a comparison depend on the length of the updating interval τ: clearly, $\tau_1 = \Delta t = 10$ min for C2t and $\tau_2 = 12$ h for C2i are extreme. It is possible that an optimal interval τ_0 lies in between, $\tau_1 < \tau_0 < \tau_2$ (e.g., Morel and Talagrand, 1974). If so, it would seem that it might be closer to τ_1 than to τ_2, e.g., $\tau_0 = 3$ h, rather than the subsynoptic interval of 6 h presently being considered by some operational groups. Further study of this question is warranted. A theoretical investigation along the lines of estimation theory is currently being undertaken (M. Ghil and J. Tavantzis, personal communication).

Of the experiments with time-continuous assimilation, DN gave a practically nil impact: $x_{DN} = 0.21$ in S_1 score, compared to $x_{CNc} = 0.97$ for the control. The statistical significance of experiment DN was also much lower than the threshold of 1; this threshold, we recall, was actually attained by CNc. The corresponding result for CN, which uses the same data, but SCM assimilation rather than DIM, is $x_{CN} = 1.83$. In fact, DN gave results similar to those of the reproducibility experiment NOr. We conclude that DIM does not properly use the data, since the same data (from Nimbus 6) can be assimilated to give a noticeable impact with a better method (SCM).

All time-continuous SCM experiments (CN, CN*l*, C2 and C2t) gave results which were at least marginally significant, $0.94 \leq \zeta \leq 2.34$, for both S_1 scores and rms errors. For the same quantity of data, CN*l* gave better results than CN, since it took into account the higher accuracy of the conventional data available over land areas. Also C2t gave impacts almost twice as high as those of C2, due to better tuning of interpolation coefficients. On the whole, the use of SCM gave quite satisfactory results, especially with data from both satellites and with properly tuned coefficients.

The effect of data quantity on the SAM sequence SV through S2b was already discussed. For the same data quantity, SNa gave results which were better than those of CN and comparable to those of CN*l*.

Experiments S2, S2a and S2b gave very similar results. This shows that present attempts at eliminating the effect of the warm bias of sounding temperatures (S2a and S2b) were not successful. The arithmetic average of the results for these three experiments appears in Table 3 on the row denoted S2m. The representative results S2m for a 2SAT SAM experiment are approximately twice as high as those for C2, and practically indistinguishable from those of C2t.

We conclude that for the same data, the assimilation method can make all the difference between obtaining negligible results (DIM), appreciable results (non-tuned SCM) or very good results (SAM or tuned SCM). It appears that both C2t and S2m give results which are at the level of a 5% improvement in mean impact for S_1 scores and of 10% in rms errors. It is known that SCM corresponds to a SAM in which the matrix A of Section 2 is diagonal (Rutherford, 1972); such diagonal approximations of banded matrices can at times yield satisfactory results.

The practical advantage of SAM is essentially that it is self-tuning. SAM can also be more easily generalized to three-dimensional and to anisotropic formulations. Such generalization is also conceivable for SCM, but the task of tuning becomes much more arduous than in an isotropic, two-dimensional version. Moreover, synoptic evaluation of forecast fields for C2t and S2a (see Section 4c) indicates that SAM is still superior to SCM.

Experiments S2μ and S2ν, which were similar in all other respects to S2a, addressed the question of relative accuracy of satellite data versus forecast values. Sounding data and forecast values were weighted in accordance with their computed variance; the forecast values also carry all the information from the other types of data, as described in Section 3. The outcome of these experiments was quite comparable to that of the other 2SAT SAM

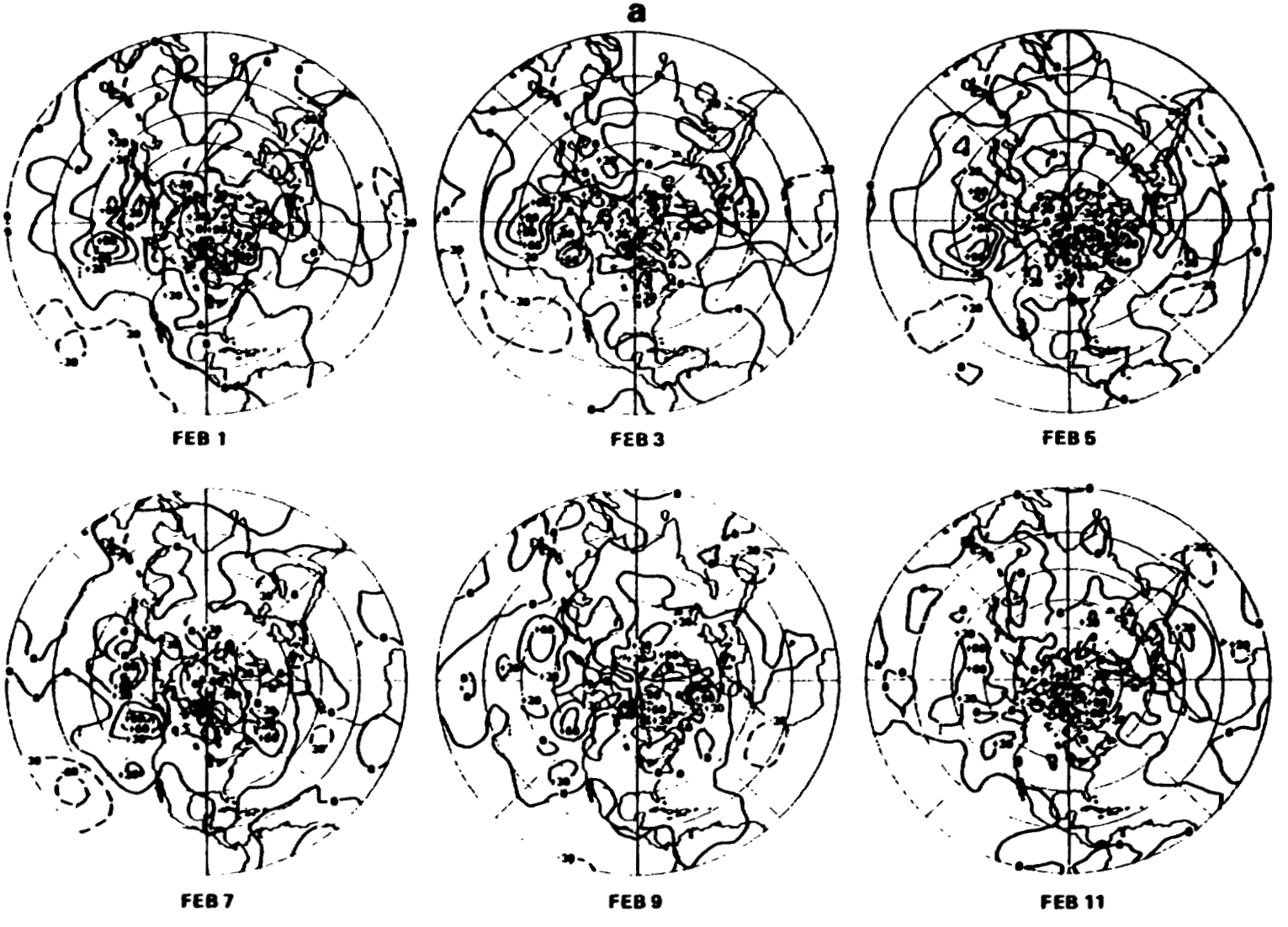


FIG. 5. Initial state differences in 500 mb geopotential heights for the Northern Hemisphere between experiments C2t (2SAT/SCM) and NO (NOSAT), $\phi_{C2t} - \phi_{NO}$. Positive differences are indicated by full contour lines, negative differences by dashed lines. The differences are given at 0000 GMT. They show the same warm bias of satellite sounding temperatures apparent in Fig. 3.

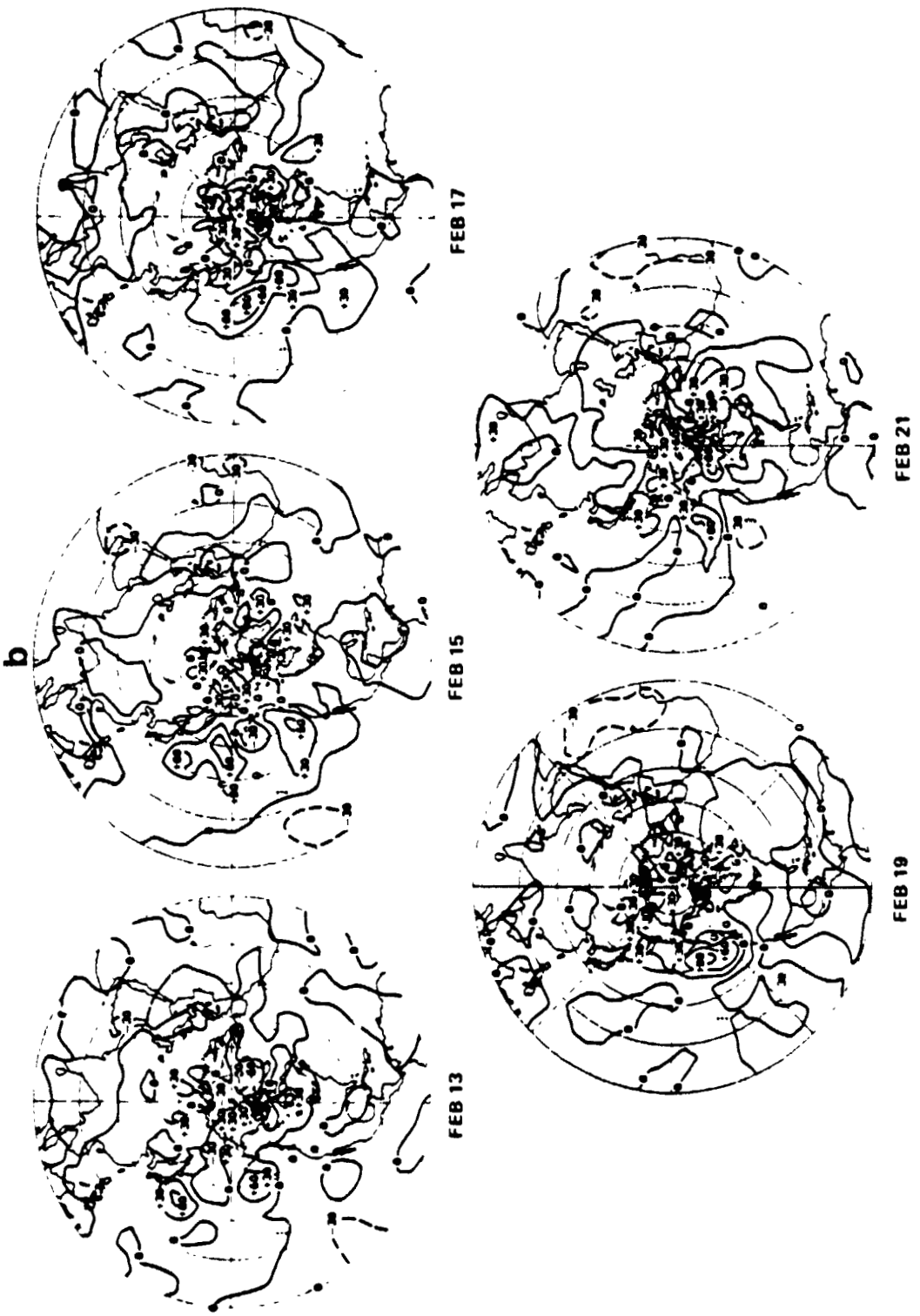


FIG. 5 *Continued*

TABLE 2. S_1 skill score impact in the 500 mb height field verified at 48 h over North America. The columns labeled 2/1–2/21 contain the results for the forecasts started at 0300 GMT on the corresponding days. The first row contains the actual skill score values for the NOSAT Experiment NO; the other rows contain the impacts determined by the differencing process described in the text. In the next six columns an overview of the results in the preceding columns appears. The procedure for deriving this overview is also described in the text.

Experiment	Day forecast started											Actual value			Difference from NO				Percent impact
	2/1	2/3	2/5	2/7	2/9	2/11	2/13	2/15	2/17	2/19	2/21	Ave	SD	SE	Ave	SD	SE	Av/SE	
NO	38.9	36.8	41.1	47.2	41.4	31.4	40.5	40.1	32.7	42.7	42.6	39.58	4.54	1.37	0.0	0.0	0.0	0.0	0.0
NOr	−0.4	0.3	0.6	−1.9	−0.6	0.7	1.0	0.0	0.2	−0.9	0.6	39.55	4.92	1.48	0.03	0.69	0.21	0.17	0.09
DN	−2.3	1.2	3.1	2.7	−1.8	0.1	5.0	0.6	−2.5	−3.2	−0.3	39.33	4.64	1.40	0.25	2.61	0.79	0.32	0.63
CNc	−1.4	−0.6	1.7	0.9	0.1	0.1	1.7	−2.2	0.9	−1.4	−1.3	39.71	4.79	1.44	−0.13	1.35	0.41	−0.31	−0.32
CN	1.3	2.6	1.1	4.4	3.7	1.3	4.7	1.6	−1.4	−2.8	0.9	38.00	4.41	1.33	1.58	2.29	0.69	2.30	4.00
CN/	0.5	2.6	−0.9	4.5	3.4	−0.1	6.1	3.2	1.1	−4.3	−0.9	38.20	5.11	1.54	1.38	2.94	0.89	1.55	3.48
C2	3.9	4.4	−0.4	4.1	5.6	1.1	3.4	3.9	−4.3	−6.9	2.0	38.05	5.38	1.62	1.53	3.95	1.19	1.29	3.87
C2t	0.5	1.9	0.8	6.0	5.7	2.2	6.9	4.3	−0.5	−4.5	0.8	37.38	5.01	1.51	2.20	3.36	1.01	2.17	5.56
C2i	−1.06	1.77	2.55	0.18	1.45	4.06	2.92	6.57	−1.13	−1.32	1.68	37.98	5.42	1.64	1.61	2.42	0.73	2.20	4.06
SV	0.7	−0.5	0.2	−0.1	2.6	−1.6	5.3	1.6	−2.3	*	0.8	38.60	4.07	1.29	0.67	2.18	0.69	0.97	1.69
SNa	1.2	1.6	0.1	4.0	5.2	−0.6	6.5	1.7	−2.6	−6.5	2.1	38.44	4.85	1.46	1.14	3.61	1.09	1.05	2.88
S2	1.5	3.1	−1.9	3.3	5.2	1.9	7.5	3.8	−2.4	−3.3	2.3	37.66	5.09	1.54	1.92	3.32	1.00	1.92	4.86
S2a	0.6	3.7	−2.0	1.3	5.1	3.9	7.3	5.9	−2.6	−3.7	1.5	37.67	5.90	1.78	1.91	3.63	1.09	1.74	4.82
S2b	1.9	3.1	−3.3	2.9	5.0	1.6	7.5	3.7	−1.6	−3.7	2.4	37.81	5.35	1.61	1.77	3.41	1.03	1.72	4.47
S2μ	2.4	2.6	0.3	0.6	3.1	3.4	6.7	3.3	−0.8	−2.0	0.5	37.76	5.47	1.65	1.82	2.42	0.73	2.50	4.61
S2ν	1.6	2.8	0.1	1.6	3.4	2.3	5.0	2.6	−1.5	−1.3	0.5	38.05	4.85	1.46	1.54	1.98	0.60	2.57	3.88

* Tape reading error; summarizing results for SV were computed omitting the value for 2/19

experiments, S2, S2a and S2b. The detailed numerical results show an improvement in forecasts at 12 and 24 h over the latter, since the initial conflict over data-dense areas between sounding information and conventional information is eliminated. At later times synoptic analysis shows that the impact of the sounding temperatures is somewhat diminished, since they are given less weight than in the preceding experiments.

To evaluate the potential significance for NWP of a 5% S_1 impact and a 10% rms impact, the results in Table 5 are shown. In the table, actual S_1 and rms values are given, averaged over the 11 forecasts made from 1 February through 21 February 1976, for the NOSAT experiment NO, verified against the NMC operational analysis. The table shows that the S_1 skill score deterioration between 48 and 72 h is about 10–16%; the corresponding figure for rms errors is about 13–24%.

We conclude that for the GLAS model and the given time of the year (Northern Hemisphere winter), the impact of the available temperature sounding data assimilated by the best methods tested was equivalent to an extension of useful numerical prediction by ~8–16 h. In other words, a 60 h forecast could be made with the accuracy of today's 48 h forecast.

It appears in general that differences in initial states correlate reasonably well with impact on forecasts from those initial states (see Section 4a). Our numerical measures of impact also seem to correlate positively with improvements in the capability of predicting local weather when using the large-scale numerically predicted fields for guidance (see Section 4d).

TABLE 3. Summary of results for the numerical measures of impact.

Experiment	Percent impact		Statistical significance (Average/standard error)	
	S_1	rms	S_1	rms
NO	0	0	—	—
NOr	0.24	7.18	0.45	2.12
DN	0.21	2.43	0.13	0.94
CNc	0.97	1.51	1.07	1.23
CN	1.83	6.31	1.01	1.82
CN/	2.98	7.36	1.36	1.24
C2	2.75	5.13	0.94	1.11
C2t	5.01	9.31	2.10	2.34
C2i	2.79	3.28	1.75	1.33
SV	1.20	3.66	0.66	1.19
SNa	3.11	5.97	1.39	1.37
S2	4.74	9.57	1.94	2.14
S2a	4.46	12.41	1.85	2.59
S2b	4.79	10.02	2.05	2.49
S2m	4.66	10.67	1.95	2.41*
S2μ	4.09	12.09	1.99	4.08
S2ν	3.38	10.52	1.83	3.72

* S2m gives the mean results for experiments S2, S2a, S2b.

c. *Synoptic evaluation of forecast differences*

It is considerably more difficult to carry out a subjective comparison of prognostic fields based on synoptic criteria than to compute a numerical measure of the difference between such fields. Therefore we have to limit ourselves at this point to the presentation of a comparison of the results for the 2SAT SAM experiment S2a with those for the NOSAT

TABLE 4. Summary of numerical results for experiments extended to the entire DST-6 period.

Experiment	Percent impact (S_1/rms)			Statistical significance: Ave/SE (S_1/rms)		
	2/1–2/21	2/23–3/4	2/1–3/4	2/1–2/21	2/23–3/4	2/1–3/4
S2a	4.46/12.41	−0.10/4.64	2.90/9.98	1.85/2.59	0.00/0.72	1.52/2.35
S2μ	4.09/12.09	1.53/5.47	3.31/10.17	1.99/4.08	0.55/0.36	2.00/3.67
S2ν	3.38/10.52	−0.45/0.92	2.06/7.42	1.83/3.72	−0.26/0.29	1.37/2.88

experiment NO. In this and the next subsection we shall again refer to Experiment S2a simply as 2SAT.

The synoptic evaluation of forecasting impact for the satellite sounding data was performed by comparing the prognostic charts generated from 2SAT and NOSAT initial conditions with each other, and with the corresponding NMC analyses for verification. Comparisons were made every 12 h for a variety of prognostic fields. The results of our comparisons for the sea-level pressure and 500 mb height fields verified over North America at the end of each 72 h forecast are presented in Table 6.

The table gives the consensus of three experienced forecasters in their subjective assessment of the prognostic charts. During this phase of the evaluation, the forecasters had no knowledge of which prognostic charts were based on 2SAT initial conditions. Following Atkins and Jones (1975), each forecast has been classified according to the following scale:

A 2SAT significantly better than NOSAT
B 2SAT better than NOSAT
C 2SAT and NOSAT of equal quality
D 2SAT worse than NOSAT
E 2SAT significantly worse than NOSAT.

TABLE 5. Numerical measures of accuracy for the NOSAT experiment NO over the period 1–21 February 1976. The first column for each heading gives the corresponding measure of accuracy (S_1 or rms) at time t; the next column gives differences occurring over 12 h, e.g., $S_1(t + 12\text{ h}) - S_1(t)$, and so on; the third column gives the differences occurring from 24–48 h and from 48–72 h. The last (boxed) entry gives either the quantity $\{S_1(72\text{ h}) - S_1(48\text{ h})\}/S_1(72\text{ h})$ or the corresponding rms quantity

Time (h)	North America						Europe					
	S_1	$\Delta_{12}S_1$	$\Delta_{24}S_1$	rms	Δ_{12}rms	Δ_{24}rms	S_1	$\Delta_{12}S_1$	$\Delta_{24}S_1$	rms	Δ_{12}rms	Δ_{24}rms
					p_s (mb)							
12	45.76			3.74			38.82			3.57		
		11.30			1.39			10 39			1.36	
24	57.06			5.13			49.21			4.93		
		7.78			1.43			4.68			1.48	
36	64 87		15.11	6.56		2.72	53.89		11.33	6.41		2.89
		7.30			1.29			6.65			1.41	
48	72.17			7.85			60.54			7.82		
		4.99			0.89			5.44			0.64	
60	77.16		8.01	8.74		1.19	65.98		11.09	8.46		2.30
		3.02			0.30			5.65			1.66	
72	80.18			9.04			71.63			10.12		
			9.99%			13.16%			15.48%			22.73%
					ϕ(m)							
12	26.42			34.09			31.05			36.32		
		4.89			11.59			7.28			8.91	
24	31.31			45.68			38.33			45.23		
		5.13			14.11			6.28			19 68	
36	36.44		7.99	59.79		27.21	44.61		13.39	64.91		29 61
		2.86			13.10			7 11			9.93	
48	39 30			72.89			51.72			74.84		
		4.15			8.47			5.39			15.45	
60	43.45		7.43	81.36		15.67	57 11		10.33	90 29		23 49
		3.28			7.20			4.94			8.04	
72	46.73			88.56			62.05			98.33		
			15 90%			17.69%			16 65%			23.89%

The regions of verification and quantities verified are defined in the text.

TABLE 6. Synopsis of subjective assessment of forecasts from the 2SAT/SAM Experiment S2a compared against NO.

a. Sea-level pressure

Category	12 h	24 h	36 h	48 h	60 h	72 h	Total
A	0	0	0	0	0	0	0
B	0	1	1	4	3	4	13
C	8	8	7	5	6	5	39
D	3	2	3	2	2	2	14
E	0	0	0	0	0	0	0

b. 500 mb geopotential height

Category	12 h	24 h	36 h	48 h	60 h	72 h	Total
A	0	0	0	0	0	2	2
B	2	2	2	3	4	4	17
C	8	9	9	5	7	4	42
D	1	0	0	3	0	1	5
E	0	0	0	0	0	0	0

This classification system is designed to represent the relative utility of the prognostic charts generated from 2SAT and NOSAT initial conditions. The term significant in categories A and E implies that major prognostic differences have occurred. Categories B and D imply that there has been only a slight alteration of the prognostic pattern.

From the table it can be seen that at sea level none of the prognostic charts has been classified as A or E and that an almost equal number of B and D marks have been assigned. Although no significant impact has occurred at this atmospheric level, a tendency for more beneficial than negative impacts at 48, 60 and 72 h is evident.

At 500 mb, two 72 h prognostic charts have been classified as A and there were no E marks. In addition more than three times as many B as D marks have been assigned. Impacts at this level are most consistently beneficial at 60 and 72 h.

One example of a beneficial forecasting impact over North America is presented in Fig. 6. Figs. 6a and 6b depict the 72 h 500 mb prognostic charts from 19 February 1976 for the NOSAT and the 2SAT case, respectively. The corresponding NMC analysis is depicted in Fig. 6c. A comparison of these three charts reveals that the NOSAT prognosis is noticeably slow in its movement of a trough into the Midwest of the United States and of a ridge off the East Coast. Both of these systems are displaced further east by the 2SAT prognosis, in better agreement with the analysis. A detailed synoptic study of the 72 h prognostic fields over North America was also carried out for the forecast started on 0300 GMT 9 February 1976 (Atlas and Sakal, 1978). It showed considerable improvement in the steering of weather systems by the 2SAT forecast.

A preliminary evaluation of C2t prognostic charts showed a considerable reduction of the beneficial impact due to satellite data for the forecast from 19 February 1976, by comparison to S2a. A number of other positive impacts in S2a also were reduced slightly in C2t. This seems to suggest that the use of SAM is still preferable to the use of SCM, even though their numerical measures of forecast accuracy were quite close.

d. Differences in local precipitation forecasts

AFM local precipitation forecasts were generated for 11 numerical forecasts, 1 February–21 February 1976, from both the 2SAT and the NOSAT prognostic fields. They were generated at 24, 48 and 72 h after initial time for 128 cities uniformly distributed over the United States. Thus a total of $2 \times 3 \times 11 \times 128 = 8448$ local precipitation forecasts was made for the comparison discussed in the sequel.

Comparisons of city precipitation forecasts for these 128 cities are presented in Table 7. The number of correct precipitation forecasts for the 2SAT and NOSAT systems are shown only for those cases in which the two systems gave rise to different forecasts. The systems actually differed in about 10% of the total number of forecasts and the results given cover the cases when one system forecasts a precipitation event, while the other one forecasts no precipitation. The number of different forecasts for each day is shown for all the 128 cities considered together in the lower horizontal row. At 24 and 48 h the margin of superiority of the 2SAT system over NOSAT was 12% of the total number of cases for which they gave a different result. At 72 h this margin was 19%. We examine in the upper row a separate list of 17 cities in the Midwest, selected because they are less influenced by coastal effects, mountains or the warmer Gulf convective systems; the AFM has difficulty accounting for such effects in its present form. In the case of these cities we see a margin of superiority of 2SAT over NOSAT of better than 2:1 in forecasting for 24 and 72 h and an even larger margin for 48 h.

To exhibit the correlation between numerical impact measures (Section 4b) and the impact on local weather forecasts we present Tables 8 and 9. Table 8 gives S_1 skill score impacts verified at 72 h over North America for each forecast separately and Table 9 gives the corresponding precipitation forecast impacts. In both tables the results are given for the control experiment CNc, as well as for S2a.

We observe large numerical impacts in both ϕ and p, for the forecasts from 3, 9, 11 and 13 February; they are all positive, and large when compared to those of the control experiment CNc. The precipita-

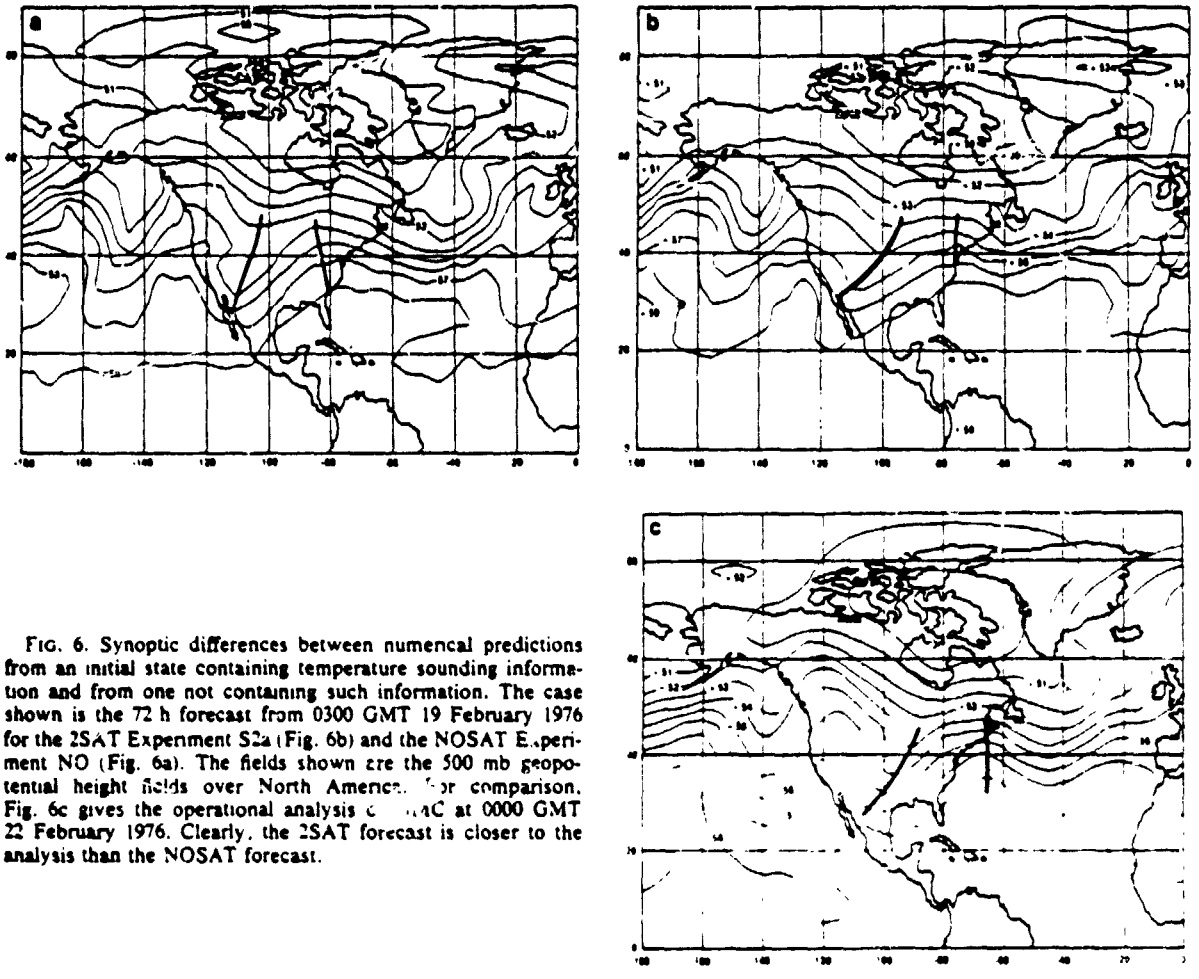


FIG. 6. Synoptic differences between numerical predictions from an initial state containing temperature sounding information and from one not containing such information. The case shown is the 72 h forecast from 0300 GMT 19 February 1976 for the 2SAT Experiment S2a (Fig. 6b) and the NOSAT Experiment NO (Fig. 6a). The fields shown are the 500 mb geopotential height fields over North America. For comparison, Fig. 6c gives the operational analysis of NMC at 0000 GMT 22 February 1976. Clearly, the 2SAT forecast is closer to the analysis than the NOSAT forecast.

tion forecast impacts at 72 h are large when compared to the NOSAT experiment for 9 and 11 February, and moderate but still positive for 3 and 15 February. A large positive impact on precipitation forecasts also occurs on 19 February; while the numerical measures did not show a strong impact of sounding data for this large-scale prognosis, the synoptic comparison (Section 4c) did show a significant improvement in the 2SAT prognostic fields.

TABLE 7. Summary of results for local precipitation forecasting in cases where use of 2SAT (S2a) and NOSAT (NO) prognostic fields produced different forecasts. "2SAT better" means that the AFM produced a correct precipitation forecast from 2SAT prognostic fields, and an incorrect one from NOSAT fields; "NOSAT better" means the opposite was the case.

	24 h		48 h		72 h	
	2SAT better	NOSAT better	2SAT better	NOSAT better	2SAT better	NOSAT better
Midwest (17 cases)	11	4	11	1	[illegible]	3
Entire U.S. (128 cases)	82	64	68	[illegible]	85	58

A large, negative impact on precipitation forecasting at 72 h occurs for 15 February when the numerical measures show little impact if at all. The control experiment CNc produced negative impacts on

TABLE 8. Daily S_1 skill score impacts verified over North America at 72 h. At the bottom of each column (the mean S impact) ± (the corresponding standard deviation) is given.

Date	500 mb heights		Sea level pressure	
	2SAT (S2a)	Control (CNc)	2SAT (S2a)	Control (CNc)
February				
1	1.8	−3.5	−4.9	−12.0
3	9.4	3.4	8.4	4.*
5	−2.2	0.8	[illegible]	−2.8
7	0.*	−1.1	10.*	0.3
9	7.8	1.2	12.*	1.4
11	5.9	−2.1	9.4	−1.2
13	6.8	1.1	[illegible]	5.2
15	1.9	−3.3	1.5	−3.*
17*	−4.5	−1.8	−5.*	−4.5
19	4.7	3.9	−2.8	5.8
21	5.4	−0.9	5.2	−4.3
Totals	1.43 ± 1.[illegible]	−0.20 ± 0.8	4.73 ± 2.3	−0.01 ± 1.6

TABLE 9. Daily results of local precipitation forecasting for cases of different forecasts being produced by the two systems. The results refer to 128 U.S. cities and are verified at 72 h.

Date	2SAT better	NOSAT better	CONTROL better	NOSAT better
February				
1	3	6	1	6
3	8	7	2	11
5	7	5	6	4
7	7	4	3	1
9	17	8	1	1
11	3	0	2	0
13	5	3	3	2
15	8	16	3	10
17	5	3	0	1
19	21	3	6	5
21	1	3	1	1
Totals	85	58	28	42

the average in both the prognostic fields and the AFM score; this reinforces our confidence in the significance of the positive results for S2a.

From these comparisons we conclude that there is a need for much more study of the interrelationship between initial state differences and different forecast impact measures. The preliminary evidence is incomplete; it seems to point, however, in the direction of positive correlation between large impacts on initial-state determination in data-sparse areas, on the one hand, and large impacts on the different quantities measuring the accuracy of forecasts at 48 and 72 h over data-rich areas, on the other.

5. Concluding remarks

Methods were developed for the time-continuous assimilation of temperature sounding data: direct insertion (DIM), asynoptic successive correction (SCM) and local linear regression (SAM). These methods were applied to DST-6 data from the operational and experimental temperature sounders aboard the NOAA 4 and Nimbus 6 satellites.

An extensive series of experiments was performed in which the satellite data were assimilated to produce initial states for numerical forecasts. For each assimilation experiment, an evenly spaced sequence of initial states was selected, from which 3 day forecasts were carried out. The experiments differed from each other by the use of various amounts of data and by the assimilation method they utilized. The effects of the satellite data were evaluated according to the following criteria: 1) differences between the initial states produced with and without utilization of satellite data, 2) differences between numerical predictions made from these initial states, and 3) differences in local weather forecasts resulting from the large-scale numerical predictions

Initial state differences were evaluated in terms of magnitude and location of large-scale differences between meteorological fields. Numerical prediction differences were evaluated in terms of S_1 skill scores and rms errors, as well as by synoptic case studies. An automated forecasting model (AFM) based on quasi-geostrophic theory and on subjective forecasting principles was developed to facilitate the objective evaluation of differences produced in local weather forecasts, especially precipitation forecasts.

The results of our experiments suggest a number of conclusions:

1) Satellite-derived temperature data can have a modest, but statistically significant positive impact on numerical weather forecasts, as verified over the continents of the Northern Hemisphere.

2) This impact is highly sensitive to the quantity of the data available—the impact of a two-satellite system is larger than that of one satellite by an amount roughly proportional to the quantity of data provided.

3) The assimilation method plays a major role in the magnitude of the impact for the same data—direct insertion had a practically null impact, while SCM and SAM provided an appreciable impact.

Experiments showed that time-continuous assimilation of remote-sounding temperatures is superior to their intermittent assimilation. The impact for the best method tested and for the full amount of data available was about 5% in S_1 score and 12% in rms errors; these results correspond to the possibility of extending by about 8–16 h the usefulness of numerical weather prediction (NWP) in the range between 48 and 72 h. There are indications that local weather forecasts using large-scale NWP results as guidance can be similarly improved.

We are in the process of further refining our statistical assimilation method for sounding data and of making it more efficient. We intend also to adapt SCM and SAM to the assimilation of cloud-track wind data from geostationary satellites, and we expect eventually to apply both methods to FGGE (First Global GARP Experiment) data.

Based on the experience from this study, we believe that there are *two* major areas in which improvements can bring about larger impacts of satellite-borne sounding systems. First, instrument development to improve the accuracy and vertical resolution of the sounder temperature profiles themselves. At present, vertical temperature profiles derived from satellite sounding radiance data have 2–2.5°C rms errors when compared with colocated radiosonde profiles. Although this accuracy falls short of meeting GARP requirements, we have shown that the data are still capable of producing

modest impacts when properly utilized. However, the deficiencies in the quality of the data can only partially be compensated for by special retrieval and assimilation methods.

There is certainly a need for continuous data monitoring and for continued development of retrieval and of assimilation methods. Moreover, close interaction between retrieval and assimilation methods seems to hold great promise. But the most important requirement is still that of instrument packages with greater accuracy and resolution, capable of providing reliable soundings under all atmospheric conditions.

Second, numerical prediction models themselves must be improved to make better use of the sounding data. In order to successfully assimilate asynoptic data and have them contribute to more accurate forecasts, it is necessary that the model be able to convey information accurately over extended distances and periods of time from one region of the globe to another. Improvement in forecasting models can only proceed by a judicious combination of higher grid resolutions, more accurate numerical discretization methods, and better representations of atmospheric processes in the model.

We hope that a concerted effort in designing better observing instruments and systems, refining the methods for processing and assimilating their observations, and developing better numerical models will lead to considerable improvements in numerical weather prediction and to a better understanding of the atmospheric circulation; these are the goals of the Global Atmospheric Research Program.

Acknowledgment. It is a pleasure to acknowledge all the contributors to the Goddard Sounding Temperature Impact Test: M. Almeida, P. Anolick, M. Eaton, M. Economedes, J. Forkosh, D. Han, M. Iredell, T. Kern, T. Kimmel, R. Kinkhabwala, J. Majewski, I. Morrow, R. Mosebach, K. Newman, L. Peng, W. J. Quirk, R. Rosenberg, N. Rushfield, C. Ryan, J. Searl, B. Shkoller, A. Smith, P. Suchanick, Y. Sud, E. Tobenfeld, L.-C. Tsang, C.-Y. Tung, E. Umland, R. van Norton and K. K. Wong. Especially helpful were H. Carus R. Dilling, D. Edelmann and D. Sakal. Discussions with M Cane, J. Susskind and H. J. Thiebaux were very useful. Typing and graphical work were performed speedily and cheerfully by C. Engle, C. Fonner, S. Mathis, D. Rosen, L. Rosen and C. Ullrich.

E. Rivas and J. Shukla read the manuscript carefully and made thoughtful comments. The referees' suggestions were also very helpful.

TABLE A1. Forecast parameters and their units.

Forecast parameter	Abbreviation	Units
Vorticity advection (absolute) at 500 mb	VA	10^{-6} s^{-2}
Temperature advection at 850 mb	TA	10^{-4} °C s^{-1}
Thickness (1000 mb–500 mb) advection	THA	10^{-5} m^2 s^{-3}
Moisture advection at 850 mb	MA	10^{-8} g g^{-1} s^{-1}
Dew-point depression at 850 mb	DD	°C
Temperature at 850 mb	T850	°C
Wind direction at sea level (geostrophic)	WD	deg
Vertical wind shear from sea level to 850 mb and from 850 to 750 mb	VWS	qualitative

Note: The last parameter is not used in the current version of the AFM; it will be used in attempting to account for low-level warm frontal precipitation.

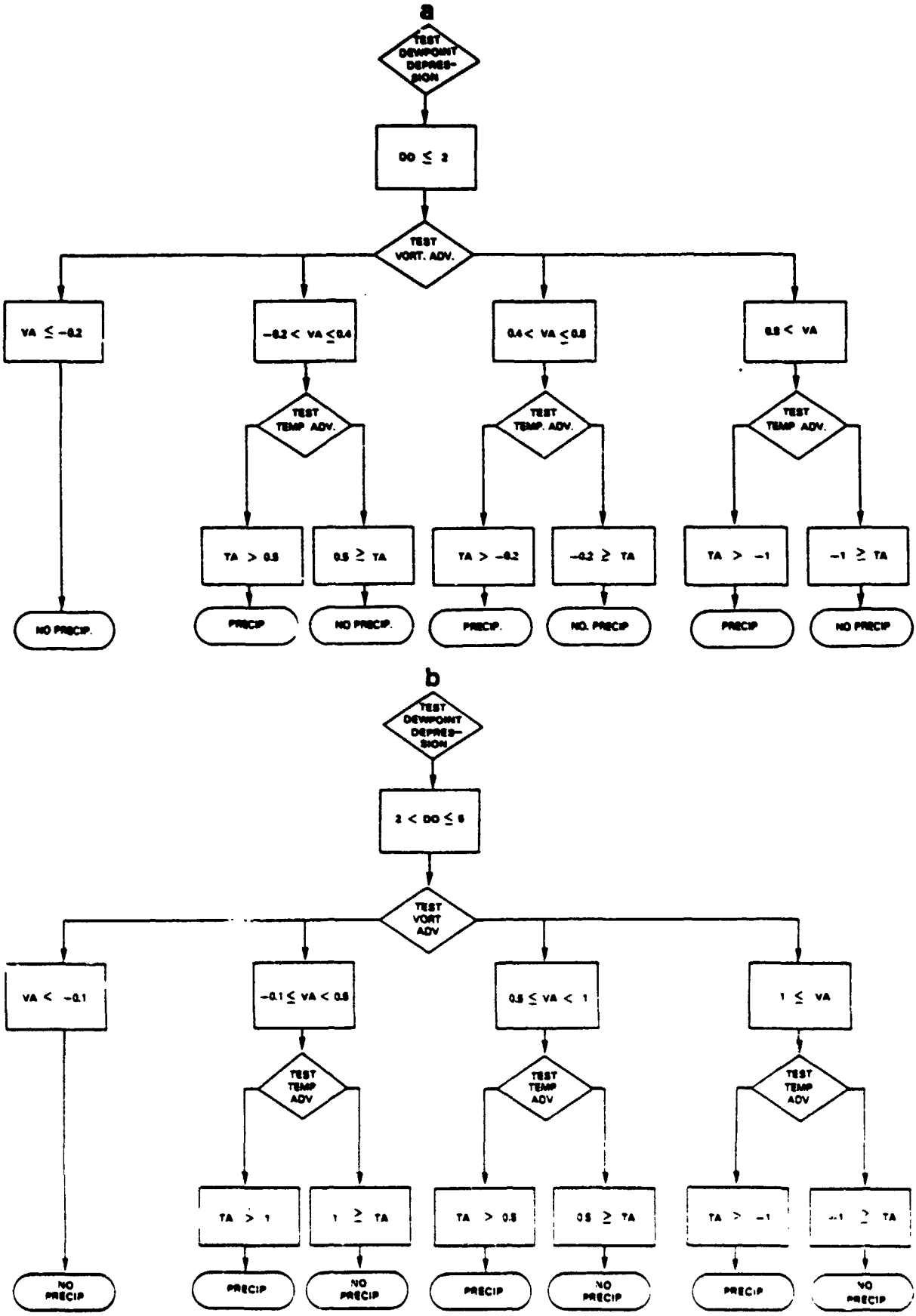
APPENDIX

AFM Flow Diagrams

The Automated Forecasting Method (AFM) is based on most of those criteria used in subjective forecasts which can be easily quantified. These empirical criteria are grounded theoretically in basic quasi-geostrophic relationships (Houghton and Irvine, 1976). The forecast parameters the AFM uses are given in Table A1, together with the appropriate units.

In the following figures, the logical structure of the AFM is described using a flow diagram presentation. Fig. A1 gives the flow diagram for determining large-scale precipitation. Fig. A2 contains the flow diagram used in deciding whether cold frontal precipitation will occur. Fig. A3 represents the flow diagram for forecasting the occurrence of local showers. The AFM will issue a precipitation forecast if either one of these types of precipitation is predicted to occur.

Fig. A4 shows the test carried out in order to determine the type of precipitation, given that precipitation does occur. Fig. A5 contains a sample computer output of the current version of the AFM


a
TEST DEWPOINT DEPRES-SION
DD ≤ 2
TEST VORT. ADV.
VA ≤ −0.2
−0.2 < VA ≤ 0.4
0.4 < VA ≤ 0.8
0.8 < VA
TEST TEMP ADV.
TEST TEMP. ADV.
TEST TEMP ADV.
TA > 0.8
0.8 ≥ TA
TA > −0.2
−0.2 ≥ TA
TA > −1
−1 ≥ TA
NO PRECIP.
PRECIP
NO PRECIP.
PRECIP.
NO. PRECIP
PRECIP
NO PRECIP
b
TEST DEWPOINT DEPRES-SION
2 < DD ≤ 6
TEST VORT ADV
VA < −0.1
−0.1 ≤ VA < 0.8
0.8 ≤ VA < 1
1 ≤ VA
TEST TEMP ADV
TEST TEMP ADV
TEST TEMP ADV
TA > 1
1 ≥ TA
TA > 0.8
0.8 ≥ TA
TA > −1
−1 ≥ TA
NO PRECIP
PRECIP
NO PRECIP
PRECIP
NO PRECIP
PRECIP
NO PRECIP

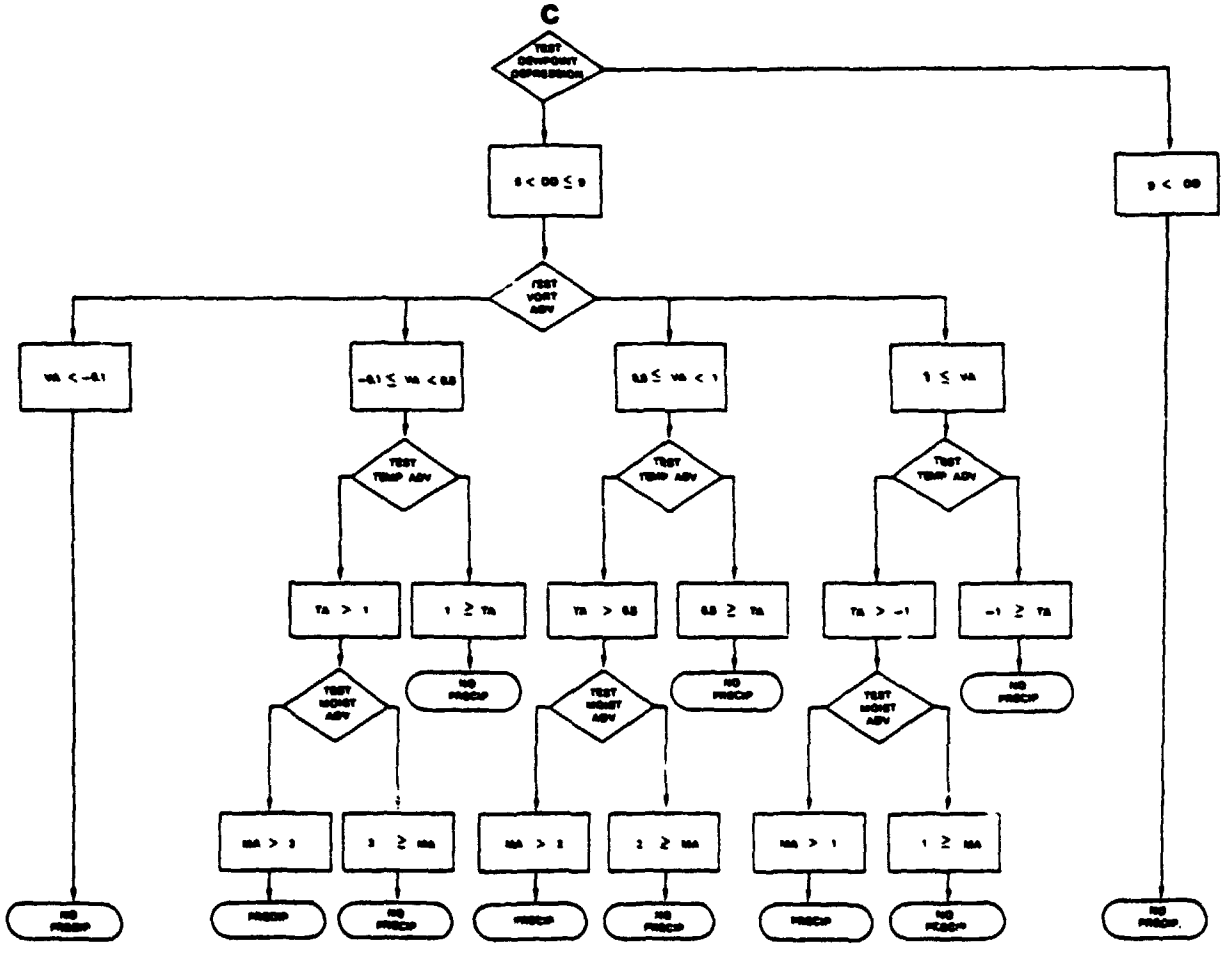


FIG. A1. Flow diagram of the procedure used to determine the occurrence of large-scale precipitation.

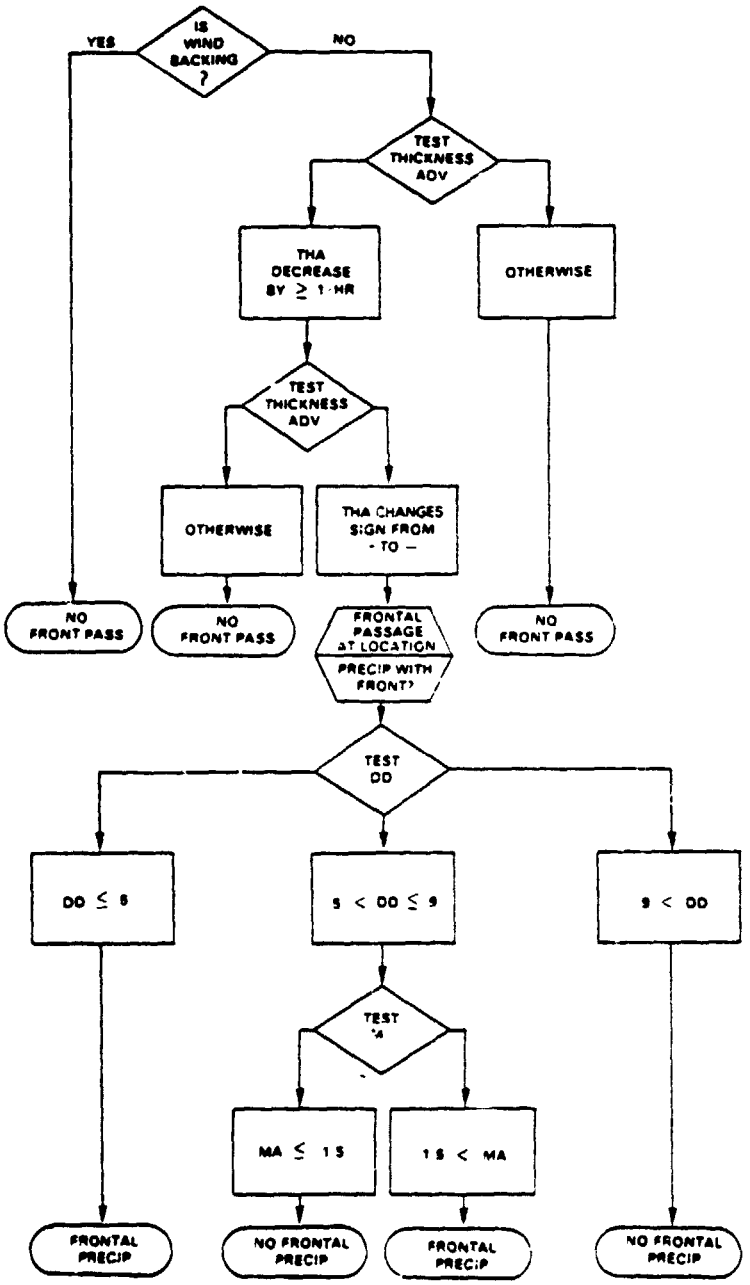


FIG. A2. Flow diagram for the occurrence of cold frontal precipitation.

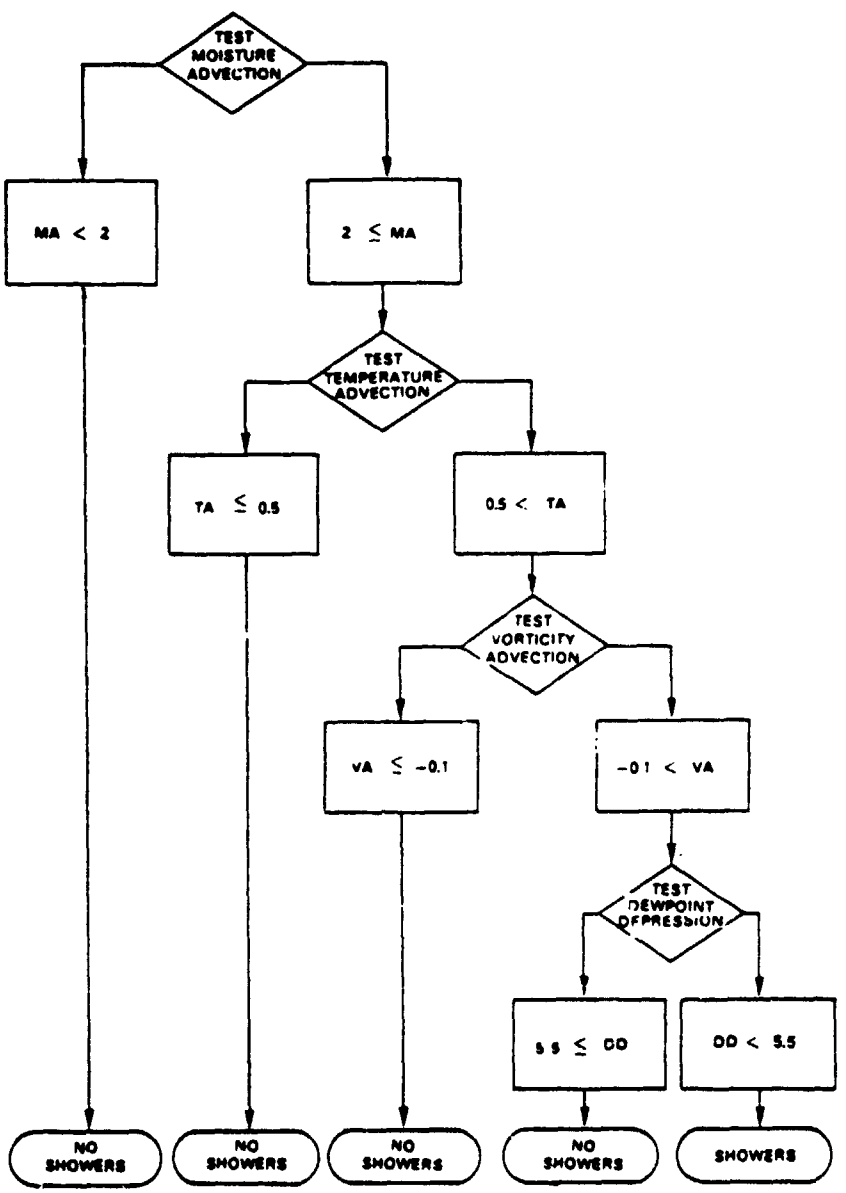


FIG. A3. Flow diagram of the procedure used to forecast the occurrence of local showers.

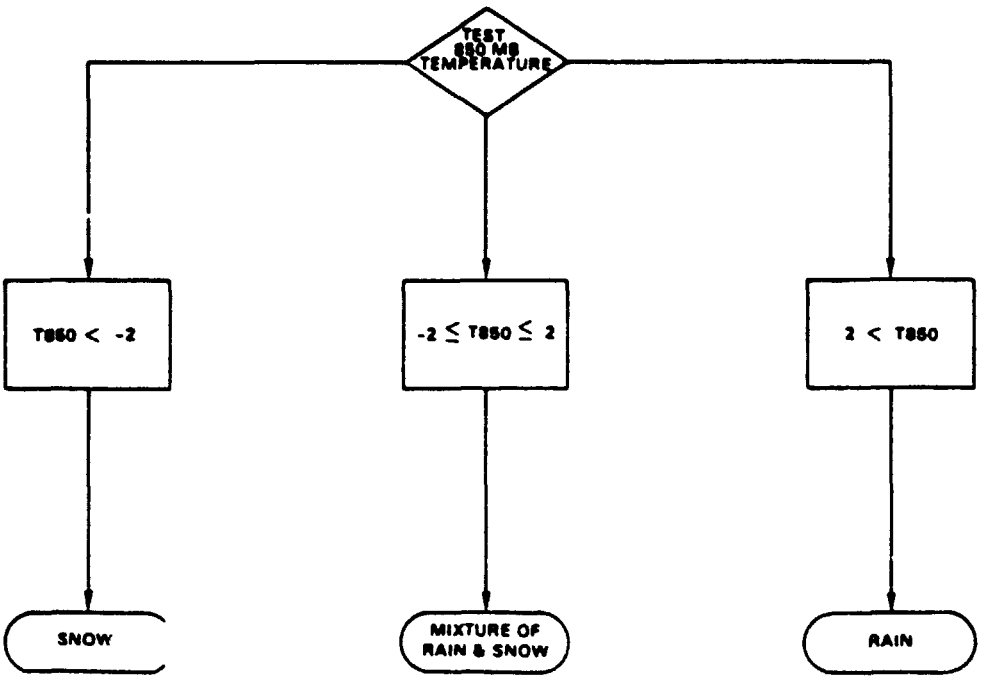


FIG. A4. Flow diagram of the procedure used to determine the type of precipitation.

24-HOUR SUMMARY

NEW YORK

2 0Z	DD= 6.11	VA= -0.31	TA= 0.61	MA= 2.96	THADV= 5.08	VORT= 10.49	T850= 0.21	U= 13.14
2 4Z	DD= 4.22	VA= -0.07	TA= 0.37	MA= 1.98	THADV= 1.26	VORT= 9.93	T850= 0.08	U= 14.81
2 8Z	DD= 2.22	VA= 0.37	TA= 0.09	MA= 1.41	THADV= -3.23	VORT= 9.66	T850= -0.30	U= 12.12
212Z	DD= 1.18	VA= 0.99	TA= -0.24	MA= 0.33	THADV= -4.18	VORT= 9.69	T850= -0.76	U= 6.43
216Z	DD= 0.31	VA= 1.28	TA= -0.62	MA= -1.03	THADV= -9.78	VORT= 10.14	T850= -1.52	U= 4.56
220Z	DD= 0.17	VA= 1.04	TA= -0.88	MA= -2.51	THADV=-21.52	VORT= 11.01	T850= -2.25	U= 5.71

12-HOUR PRECIPITATION FORECAST

2 0Z	SHOWERS
212Z	RAIN & SNOW

FRONT FORECAST

NO FRONT PASSED	22Z +/- 2 HRS	
FRONT PASSED	2Z +/- 2 HRS	PRECIPITATION
NO FRONT PASSED	6Z +/- 2 HRS	
NO FRONT PASSED	10Z +/- 2 HRS	
NO FRONT PASSED	14Z +/- 2 HRS	
NO FRONT PASSED	18Z +/- 2 HRS	

FIG. A5. Sample computer output of the automated local precipitation forecasting method.

REFERENCES

Arakawa, A., and V. R. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods in Computational Physics*, Vol. 17, J. Chang, Ed., Academic Press, 173–265.

Atkins, M. J., and M. V. Jones, 1975: An experiment to determine the value of satellite infrared spectrometer (SIRS) data in numerical forecasting. *Meteor. Mag.*, **104**, 125–142.

Atlas, R., and D. Sakal, 1978: Evaluation and verification tests. Chap. 5 of Halem *et al.* (1978), 5.1–5.83.

Automation Division Staff 1973: NMC format for observational data (ADP Reports). National Meteorological Center Office Note 29, Rev. 2. NOAA/National Weather Service, 50 pp.

Bengtsson, L., 1975: Four-dimensional data assimilation of meteorological observations. *GARP Publ. Ser.*, No. 15, WMO/ICSU, Geneva, Switzerland, 76 pp.

——, and N. Gustavsson, 1971: An experiment in the assimilation of data in dynamical analysis. *Tellus*, **23**, 328–336.

——, and ——, 1972: Assimilation of non-synoptic observations. *Tellus*, **24**, 383–399.

Cane, M., V. Cardone, M. Halem, I. Halberstam and J. Ulrich, 1978: Realistic simulations of the global observation system of Seasat-A marine wind data. Submitted to *Mon. Wea. Rev.*.

Charney, J., M. Halem and R. Jastrow, 1969: Use of incomplete historical data to infer the present state of the atmosphere. *J. Atmos. Sci.*, **26**, 1160–1163.

Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **85**, 367–374.

Gandin, L. S., 1963: *Objective Analysis of Meteorological Fields*. Gidrometeor. Izd., Leningrad. English translation by Israel Program for Scientific Translations, Jerusalem, 1965. [NTIS N6618047, Library of Congress QC996.G3313.]

Ghil, M., 1975: Initialization by compatible balancing. Rep. 75-16, Inst. Comp. Appl. Sci. Eng., NASA Langley Research Center, 38 pp.

——, and R. Dilling, 1978: Analysis and assimilation. Chap. 3 of Halem *et al.* (1978), 3.1–3.173.

——, and R. Mosebach, 1978: Asynoptic variational method for satellite data assimilation. Halem *et al.* (1978), 3.32–3.49.

——, B. Shkoller and V. Yangarber, 1977a: A balanced diagnostic system compatible with a barotropic prognostic model. *Mon. Wea. Rev.*, **105**, 1223–1238.

——, R. Dilling and H. Carus, 1977b: A statistical method for the time-continuous assimilation of satellite-derived temperatures. *Preprints 5th Conf. Probability Statistics Atmospheric Sciences*, Las Vegas, Amer. Meteor. Soc., 320–324.

Halem, M., M. Ghil, R. Atlas, J. Susskind and W. J. Quirk, 1978: The GISS sounding temperature impact test. NASA Tech. Memo. 78063, Goddard Space Flight Center, 421 pp. [NTIS N7831567].

Hayden, C. M., 1973: Experiments in the four-dimensional assimilation of Nimbus-4 SIRS data. *J. Appl. Meteor.*, **12**, 425–436.

Herman, G. F., and W. T. Johnson, 1978: The sensitivity of the general circulation to Arctic sea-ice boundaries: A numerical experiment. *Mon. Wea. Rev.*, **106**, 1649–1664.

Holton, J. R., 1972: *An Introduction to Dynamic Meteorology*. Academic Press, 319 pp.

Houghton, D. D., and W. S. Irvine, 1976: A case study comparison of the performance of operational prediction models used in the United States. *Mon. Wea. Rev.*, **104**, 817–827.

Jastrow, R., and M. Halem, 1973: Simulation studies and the design of the first GARP global experiment. *Bull. Amer. Meteor. Soc.*, **54**, 13–21.

Kasahara, A., 1972: Simulation experiments for meteorological observing systems for GARP. *Bull. Amer. Meteor. Soc.*, **53**, 252–264.

Kistler, R. E., and R. D. McPherson, 1975: On the use of a local wind correction technique in four-dimensional data assimilation. *Mon. Wea. Rev.*, **103**, 445–449.

McPherson, R. D., 1975: Progress, problems, and prospects in meteorological data assimilation. *Bull. Amer. Meteor. Soc.*, **56**, 1154–1166.

Morel, P., and O. Talagrand, 1974: Dynamic approach to meteorological data assimilation. *Tellus*, **26**, 334–344.

Murphy, A., and D. Williamson, Coordinators, 1976: *Weather Forecasting and Weather Forecasts: Models, Systems, and Users*. Notes from a Colloquium held at the National Center for Atmospheric Research, Summer 1976, NCAR/CQ-5 + 1976-ASP, Vols. 1 and 2, 900 pp. [NTIS PB-268 913/914].

Peng, L., and B. Shkoller, 1978: Filtered equations method. Halem *et al.* (1978), 3.50–3.88.

Phillips, N., 1976: The impact of synoptic observing and analysis systems on flow pattern forecast. *Bull. Amer Meteor. Soc.*, **57**, 1225–1240.

Rutherford, I. D., 1972: Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.*, **29**, 809–815.

——, 1973: Experiments in the updating of P. E. forecasts with real wind and geopotential data. *Preprints 3rd Conf. Probability Statistics Atmospheric Sciences*, Boulder, Amer. Meteor. Soc., 198–201.

Schlatter, T. W., G. W. Branstator and L. G. Thiel, 1977: Reply (to Comments by H. J. Thiebaux on "Testing a global multivariate statistical objective analysis scheme with observed data). *Mon. Wea. Rev.*, **105**, 1465–1468.

Smith, W. L., and H. M. Woolf, 1976: The use of eigenvectors of statistical covariance matrices for interpreting satellite sounding radiometer observations. *J. Atmos. Sci.*, **33**, 1127–1140.

Somerville, R. C. J., P. H. Stone, M. Halem, J. E. Hansen, J. S. Hogan, L. M. Druyan, G. Russell, A. S. Lacis, W. J. Quirk and J. Tenenbaum, 1974: The GISS model of the global atmosphere. *J. Atmos. Sci.*, **31**, 84–117.

Stone, P. H., S. Chow and W. J. Quirk, 1977: The July climate and a comparison of the January and July climates simulated by the GISS general circulation model. *Mon. Wea. Rev.*, **105**, 170–194.

Susskind, J., and D. Edelmann, 1978: Sounding temperature studies. Chap. 2 of Halem *et al.* (1978), 2.1–2.82.

Thiebaux, H. J., 1977: Extending estimation accuracy with anisotropic interpolation. *Mon. Wea. Rev.*, **105**, 691–699.

Werbowetzki, A., 1975: Indirect sounding of the atmosphere from NOAA spacecraft–regression after categorization, method and results. *Preprints 4th Conf Probability Statistics Atmospheric Sciences*, Tallahassee, Amer. Meteor. Soc., 165–170.

Paper 12

NASA Conference Publication 2076 – Fourth NASA Weather and Climate Program Science Review, January 24-25, 1979, NASA/GSFC, Greenbelt, MD.

Paper No. 55

SOME ASPECTS OF HYDROLOGY IN THE GLAS GCM

R. Godbole, *National Research Council, Washington, DC and Goddard Space Flight Center, Greenbelt, Maryland*

ABSTRACT

The response of hydrology in the GLAS GCM has been evaluated. The results show that the distribution of precipitation agrees fairly well with observations and that the model tends to maintain the hydrological balance.

Paper 13

Reprinted from: *Remote Sensing of the Atmosphere: Inversion Methods and Applications*
Edited by A.L. Fymat and V.E. Zuev

SOME EXPERIMENTS ON THE EFFECT OF REMOTE SOUNDING TEMPERATURES UPON WEATHER FORECASTING

M. Halem, M. Ghil* and R. Atlas**
Laboratory for Atmospheric Science, Goddard Space Flight Center
Greenbelt, Maryland 20771, USA

ABSTRACT

We present a time-continuous statistical method for the four-dimensional assimilation of remote sounding temperatures based on radiance measurements from polar-orbiting satellites. This method is applied to DST-6 data from the NOAA-4 and Nimbus-6 satellites.

We report on experiments in which the state of the atmosphere throughout the test period was determined using a varying amount of satellite data and in which different methods were used for their assimilation. Data from the NOAA-4 satellite only, from Nimbus-6 only, and from both satellites together were used; the methods tested include different variations of our statistical method, as well as more traditional methods.

Using some of the atmospheric states thus determined as initial states, a number of eleven 72-hour forecasts was carried out for each experiment. The effect of the satellite data was studied using the following criteria: (i) differences between the initial states produced with and without utilization of satellite data, their magnitude and location, (ii) differences in the numerical predictions made from these initial states, evaluated numerically, by S_1 skill scores and RMS errors, as well as synoptically, and (iii) differences in local precipitation forecasts resulting from the large-scale numerical predictions. To facilitate the evaluation of local weather forecast differences, we developed a computerized forecasting model for precipitation.

Our conclusions from the study are that: (1) satellite-derived temperature data can have a modest, but statistically significant positive impact on numerical weather prediction in the two-to-three

*Permanent Affiliation: Courant Institute of Mathematical Sciences, New York University, New York, N. Y. 10012
**Permanent Affiliation: Department of Mechanical Engineering, State University of New York at Stony Brook, Stony Brook, N. Y. 11794.

day range; (2) this impact is highly sensitive to the quantity of data available; and (3) the assimilation method plays a major role in the magnitude of the impact for the same data. Our results suggest that larger impacts of satellite-borne sounding systems on weather forecasting could be brought about by improvements in two major areas: (a) improved accuracy and vertical resolution of sounder temperature profiles, and (b) improved numerical models for assimilation and forecasting.

INTRODUCTION

Temperature data obtained from atmospheric remote sounding by satellites have been in operational use for weather forecasting since December 1972. At that time the U.S. National Meteorological Center (NMC) began using in its daily weather analyses and forecasts the sounding temperature profiles obtained from radiance measurements made by the Vertical Temperature Profile Radiometer (VTPR) aboard the NOAA-2 polar-orbiting satellite. Since then significant improvements have been made in the technological capabilities of instruments and in the methods for deriving, or retrieving, temperature profiles from radiance measurements; a number of studies have also been made on the methods for utilizing these derived temperatures in numerical weather prediction (NWP).

Starting with the pioneering works of Charney et al. (1969) and Smagorinski et al. (1970), a continuing program for assessing the effect of using remote sounding temperatures upon NWP has been pursued at a number of research centers (Atkins and Jones, 1975; Hayden, 1973). The findings of different investigators concerning this effect, its nature and magnitude, have elicited a certain amount of controversy in the meteorological community and a definitive assessment is still to be made. This state of affairs can be attributed to a number of factors, among which the rapid development during the intervening period in all the elements of the problem: measurement characteristics, temperature retrieval methods, and utilization methods of the retrieved temperatures for NWP. Moreover, the numerical experiments set up to determine the forecast impact of sounding temperatures varied widely in their assumptions, their data and their criteria of assessment.

At present, a certain maturization has occurred in the field of remote sounding, as well as in that of impact studies, and the possibility of a partial consensus seems to be closer. We do not claim to be able to present such a consensus at this point; we hope,

however, that the impact study we present is a step in the right direction.

The meteorological group at Goddard Institute for Space Studies (GISS) has been involved for a number of years in the utilization of sounding temperatures for NWP. Most recently, during the period April 1976 - April 1977 an intensive impact study has been carried out by the group (Halem et al., 1977); we shall report here on the results of this study.

The temperature data used in the study are in a sense standard: they were provided by the U.S. National Environmental Satellite Service (NESS). The data had been obtained by the operational retrieval methods documented by NESS (Smith and Woolf, 1976; Werbowetzki, 1975) from the radiance measurement performed by the VTPR instrument aboard NOAA-4 and the HIRS and SCAMS instruments aboard Nimbus-6, during the Data System Test (DST) period DST-6 (January - March 1976). Different retrieval methods have been under development at GISS as well (Susskind and Edelmann, 1977), but their results will not be discussed here.

The methodological contribution of this study consists in the way we utilize the data and in some of the criteria we use for assessing the impact. The utilization method we propose falls in the broad category of statistical methods; these methods are currently gaining greater recognition and wider acceptance in NWP for the purpose of incorporating, or assimilating, different types of observational data into the determination, or analysis, of the initial states, from which numerical forecasts are made. Our method differs, however, from those in use or being developed at other centers: we believe it is especially useful for the assimilation of remote sounding data.

The impact of the sounding data was assessed by some standard numerical criteria in use, such as RMS errors and S_1 skill scores of the forecasts over certain large, data-dense regions. In addition, criteria which have direct bearing on the practical usefulness of the forecasts were developed and applied.

The assimilation method is succinctly developed in Section 2. Numerical experiments completed during the impact study are described in Section 3. The results of the experiments are presented and discussed in Section 4. Our conclusions follow in Section 5.

THE ASSIMILATION METHOD

Introduction

A major effort is now under way at a number of research centers in numerical weather prediction to apply statistical methods to the four-dimensional (4-D) assimilation of temperature data obtained from radiometric satellite observations. The use of statistical methods is strongly suggested by the large quantity and poor quality of the data.

At Goddard Institute for Space Studies (GISS), we have developed a statistical assimilation method (SAM) which operates in a time-continuous manner, along a sub-satellite track. This time-continuous, fully four-dimensional approach is the main feature which distinguishes our method from other attempts, which group all satellite data in 6 to 12 h "windows", and then apply the statistical techniques to the satellite data, as well as to conventional data, in synoptic fashion.

The use of cumulative statistical information on observational errors in order to compensate for the deficiencies in the amount and accuracy of the observations has led to the application of statistical methods in the objective analysis of conventional synoptic data (Eliassen, 1954; Gandin, 1963). This application became known in the meteorological literature as "optimal interpolation". The inclusion of forecast information into the specification of an initial state by statistical methods lead to "optimal interpolation" of the differences between observations and forecast values, rather than of the synoptic observations themselves. This approach has been advocated and implemented by Rutherford (1972, 1973).

Application of statistical methods not only to the objective analysis of conventional synoptic observations, but also to the time-continuous assimilation of asynoptic satellite-derived data, has been carried out for a non-divergent barotropic model by Bengtsson and Gustavsson (1971, 1972). The combination of all the ideas above led us to study the blending of satellite data in a time-continuous manner into assimilation runs of the GISS General Circulation Model (GCM, Somerville et al., 1974), by the procedure of local "optimal interpolation" of observed-minus-forecast values; the statistically determined corrections are then added at each model time step of the assimilation cycle to the forecast values.

This statistical method was compared in some of our experiments with other methods, of a more traditional nature, to wit: direct insertion (Bengtsson, 1975, p. 24), and successive correction

(Cressman, 1959). The detailed asynoptic implementation of these methods in our experiments is described in Ghil and Dilling (1977).

Description of the Statistical Method

Vertical temperature profiles obtained from satellite-based radiance measurements are grouped by 10-minute time intervals. A plot of a typical group of temperature data obtained in a 10-minute interval is shown in Figure 1. In the sequel we shall refer for the sake of brevity to satellite-derived temperatures as observed temperatures. In the present implementation of the method, only information at the same mandatory pressure level is used, i.e., the "optimal interpolation" is two-dimensional.

To simplify notation, we use a single subscript to indicate location, thus: $\underset{\sim}{k} = (i,j)$, where i stands for discretized longitude and j for discretized latitude. Let $\underset{\sim}{k}$ be an observation point (on a fixed mandatory pressure level), $T^0_{\underset{\sim}{k}}$ the observed temperature, and $T^f_{\underset{\sim}{k}}$ the (interpolated) model temperature at the observation point $\underset{\sim}{k}$. Let $\gamma_{\underset{\sim}{k}}$ be the difference between the observed and the forecast temperatures at $\underset{\sim}{k}$, $\gamma_{\underset{\sim}{k}} = T^0_{\underset{\sim}{k}} - T^f_{\underset{\sim}{k}}$. We wish to compute corrections $\delta_{\underset{\sim}{\ell}}$ to forecast values at grid points $\underset{\sim}{\ell} = (m,n)$ near the observation points $\underset{\sim}{k}$, where $\underset{\sim}{k}$ ranges over a group, or "patch", of observations such as the one shown in Figure 1. In the figure, observation points $\underset{\sim}{k}$ are marked by T's; grid points $\underset{\sim}{\ell}$ which are affected by the observations after corrections are made appear as +'s. More precisely, after applying the corrections provided by our method, the temperature at '+'-points, as well as at 'T'-points, will be different from the forecast value.

The corrections $\delta_{\underset{\sim}{\ell}}$ are computed by a linear formula,

$$\delta_{\underset{\sim}{\ell}} = \sum_{\underset{\sim}{k}} \alpha^{\underset{\sim}{k}}_{\underset{\sim}{\ell}} \gamma_{\underset{\sim}{k}} \; ; \qquad (1)$$

here and in the sequel we drop the vector notation for $\underset{\sim}{k}$ and $\underset{\sim}{\ell}$, the multi-index character of k and ℓ being tacitly understood. The coefficients $\alpha^{\underset{\sim}{k}}_{\underset{\sim}{\ell}}$ in the Equation (1) are defined as the solution of the system of linear equations

C-3

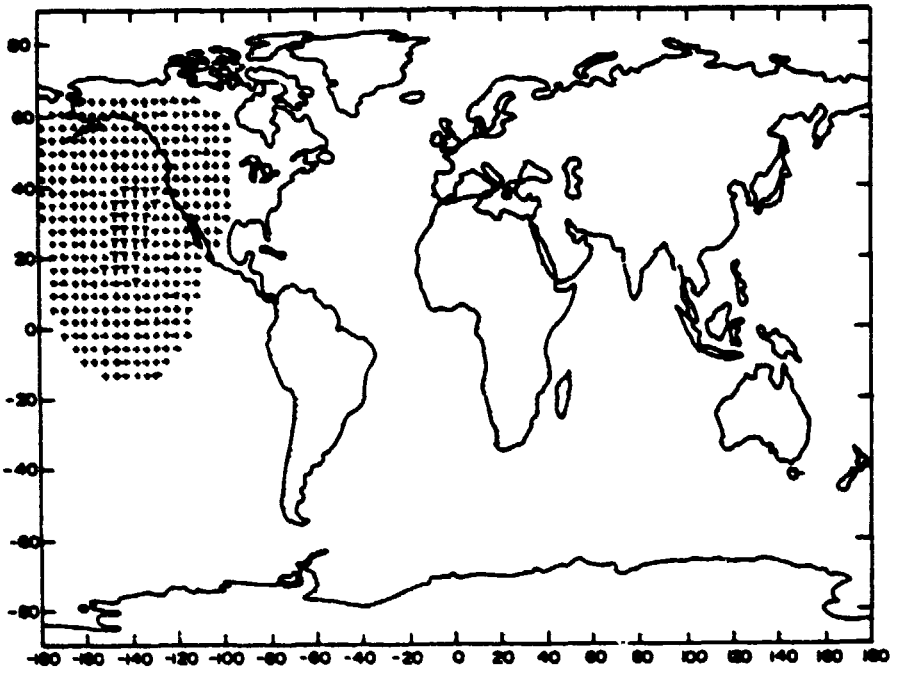


Figure 1. The distribution of temperature sounding data available over a 10 minute time interval, and of the model grid points at which our statistical assimilation method will modify forecast temperatures using these data. The 'T'-s stand for grid points in the immediate neighborhood of satellite retrievals, the '+'-s for the other points affected by the retrievals in SAM. The total domain of influence of the retrievals, consisting of 'T'-points and of '+'-points, has a radius of 2000-3000 km.

$$\underset{\sim}{A}_{\ell}\underset{\sim}{a}_{\ell} = \underset{\sim}{\beta}_{\ell} \qquad (2)$$

where $\underset{\sim}{a}_{\ell} = (a_{\ell}^{1},\ldots,a_{\ell}^{N})$, and N is the number of observation points in the patch. System (2) is the familiar normal system which arises in all statistical applications based on linear regression.

Statistical information accumulated on observations and forecasts is incorporated into the entries $a_{kk'}^{\ell}$ of the matrix $\underset{\sim}{A}$ and into the components β_{ℓ}^{k} of the right-hand side vector $\underset{\sim}{\beta}_{\ell}$. These are given by

$$a^{\ell}_{kk'} = \phi(s_{kk'}) \ , \quad \beta^{k}_{\ell} = \phi(s_{k\ell}) \ ; \tag{3}$$

here $s_{kk'}$ is the (spherical) distance between the two observation points k and k', while $s_{k\ell}$ is the distance between the observation point k and the grid point ℓ, at which we wish to make the correction. Notice that $\underset{\sim}{A}$ depends on observation points only, and we can drop the subscript ℓ, while $\underset{\sim}{\beta}_{\ell}$ depends on both observation points and the correction point we consider.

$\phi(s)$ is a correlation function. The fact that it depends only on the distance s reflects the assumptions of <u>homogeneity</u> and of <u>isotropy</u> we made at the outset concerning the error structure. The function $\phi(s)$ was computed for discrete values of s, s_1 = 100km, s_2 = 300km,..., s_p = (2p-1)x100km. The continuous function $\phi(s)$ used in (3) was obtained by fitting an analytic function $\phi = \phi(s;s_0,c)$, depending on the parameters s_0 and c, to the values $\phi_p = \phi(s_p)$, by a least-squares fit. In other words, we obtained the values of s_0 and of c for which $\phi(s;s_0,c)$ satisfied

$$\sum_p [\phi(s_p;s_0,c) - \phi_p]^2 = \text{minimum} \tag{4}$$

The actual form of ϕ used in the experiments we report on here was exponential,

$$\phi(s;s_0,c) = (1-c)e^{-s/s_0} + c \ . \tag{5}$$

The method has already been presented in Ghil <u>et al</u>. (1977b). For a more detailed description we have to refer to Ghil and Dilling (1977).

DESCRIPTION OF THE EXPERIMENTS

Our purpose was to study the effect of satellite data on the quality of initial states obtained with the aid of such data, and on the accuracy of forecasts starting from such initial states. Specifically we studied the effect of the quantity and

accuracy of the satellite data themselves on the one hand, and of the assimilation methods used to extract the information from the data, on the other.

All experiments consisted basically of continuous assimilation runs, extending over the entire period for which data were available, and of forecasts started from selected initial states produced by the assimilation run. The experiments differed from each other by the sounding data which were assimilated, and by the method which was used to carry out the assimilation. The assimilation methods used for the satellite data were direct insertion (DIM), asynoptic successive correction (SCM), and time-continuous local statistical assimilation (SAM).

The methods were applied to temperature sounding data from the NOAA-4 and Nimbus-6 satellites during the DST-6 (January-March 1976) period. A complete assimilation cycle was carried out from 00Z Jan. 29 to 03Z Feb. 21, 1976. The temperature data were supplied by the National Environmental Satellite Service (NESS). Eleven forecasts were carried out, starting at 03Z on Feb. 1,3,5,...,21. The reason to start the forecasts at 03Z and not at 00Z was to achieve as close a parallel as possible with operational practice at the National Meteorological Center (NMC) because NMC uses intermittent assimilation with a ±3h "window" for satellite data. Thus, an operational NMC forecast started at 00Z uses all asynoptic information up to 03Z; so do our experimental runs started at 03Z as we rely on time-continuous assimilation. Starting 3h after the insertion of synoptic data also has the effect of letting the model smooth out the initialization shock occurring at synoptic time (see Ghil, 1975; and Ghil et al., 1977a).

A summary description of experiments is given in Table I. A latitude-scaled correlation function (cf. Ghil and Dilling, 1977), the averaging of NESS data to GISS grid points (id.), and geostrophic wind corrections (ibid.) were used in all SAM experiments. Geostrophic wind correction was used in the DIM and SCM experiments (8240 through 8447) as well.

Experiment 8405 was set up as a control experiment to check the information content of satellite data, and to study the possible smoothing effect of SCM and SAM on the field values produced by assimilation runs. In it the temperature data used as "observations" were model values produced from a NOSAT (see Table I) 12h forecast, to which simulated observed-minus-forecast differences were added; these differences were computed at the true Nimbus-6 observation

Table I. Summary description of impact experiments. The use of data from the VTPR instrument on board the NOAA-4 satellite is denoted by VTPR, the use of the data from the HIRS and SCAMS instruments on board the Nimbus-6 satellite by NIMB. The method by which the data have been assimilated is indicated by the acronyms DIM, SCM or SAM. The experiment in which only conventional synoptic data, and no satellite data at all were used appears as NOSAT. The first digit of the code number refers to the method: DIM = 1, SCM = 2, and SAM = 3; the second digit refers to data source: VTPR = 1, NIMB = 2, and VTPR + NIMB = 3. Additional explanations concerning the method are given in the text and in Ghil and Dilling (1977).

Experiment No.	Code	Assimilation Method	Data VTPR	Data NIMB	Correlation Parameters s_0(km)	Correlation Parameters c
7578	00	NOSAT	0	0		
8240	12	DIM		X		
8405	22C	SCM*		X		
8352	22	SCM		X		
8310	23	SCM	X	X		
8447	22N	SCM**	X	X		
8574	31	SAM	X		2293	-.124
8472	32	SAM[a]		X	1831	.018
8545	33	SAM	X	X	1842	.015
8566	33a	SAM[a]	X	X	1842	.015
8593	33b	SAM[b]	X	X	1842	.015

* Control experiment with simulated Nimbus-6 data, based on NOSAT 12h forecast.
** No insertion of satellite data over data-dense land areas.
[a] Bias removed at observation point.
[b] Bias removed at grid point.

locations, by using a random number generator function with statistical properties determined from actual data.

Timing estimates for different assimilation methods show that the computational cost of implementing more sophisticated assimilation methods is not prohibitive. For example, a 24h forecast takes 40 minutes of CPU time; a 24h NOSAT assimilation runs in 48 minutes; a 24h SCM assimilation runs in 59 minutes; and a 24h SAM runs in 96 minutes (cf. Ghil and Dilling, 1977, where details are provided). These estimates all refer to the present version of the GISS GCM running on an Amdahl 470/V6 computer with a core memory of 2 megabytes. The computational cost can also be further reduced by using more efficient numerical methods and programming techniques.

DISCUSSION OF RESULTS

The impact of satellite sounding data on weather forecast has traditionally been expressed in terms of standard numerical measures such as the S_1 skill score and RMS error. However in recent years a large number of meteorologists have commented on the need for evaluating numerical forecasts in terms of the various dynamical quantitie which are used by local weather forecasters to arrive at actual weathe forecasts (Murphy and Williamson, 1976, pp. 698-900). In response to this need we attempted to develop a program which evaluates the impact of satellite data on short- and medium-range weather forecasts in the context of operational forecasting requirements. In this regard, emphasis has been placed on the practical utility of the satellite data, specifically its usefulness to the local weather forecaster

We have evaluated the impact of satellite data with the following three criteria: (1) initial state differences between analyses produced with and without the utilization of satellite data, referred to as the SAT and NOSAT initial conditions, respectively, (2) differences in the numerical predictions made from these initial conditions, referred to as the SAT and NOSAT prognostic fields, and (3) differences in the actual weather forecasts which a local forecaster would produce based on acceptance of the information given by the SAT or NOSAT prognostic fields. Differences in the prognostic fields were measured by the standard numerical criteria of S_1 skill scores and RMS errors, as well as by subjective comparisons based on synoptic criteria.

Initial State Differences

In the present context, utilization of satellite data only affects the forecasts as a result of differences between the initial states computed with (SAT) and without (NOSAT) the aid of these data. Different initial states lead then to different forecasts. We start therefore by considering the differences in initial states produced by the use of satellite data. Initial state differences are presented between experiment 8566 and the NOSAT experiment 7578. We recall that in experiment 8566 data from both the NOAA-4 and Nimbus-6 satellites were assimilated by SAM (code 33a in Table I).

Figures 2a and 2b show the 500-mb geopotential height differenc fields for the two assimilation cycles, one including all the sounding

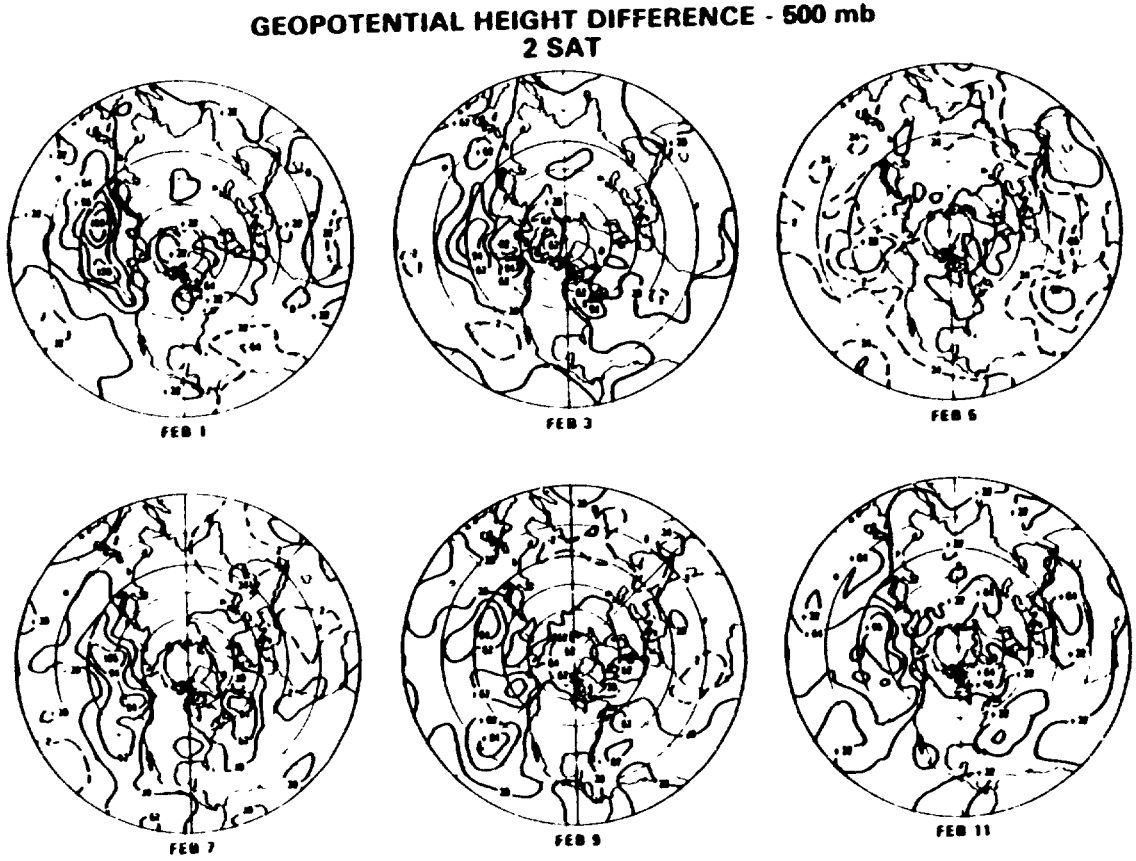
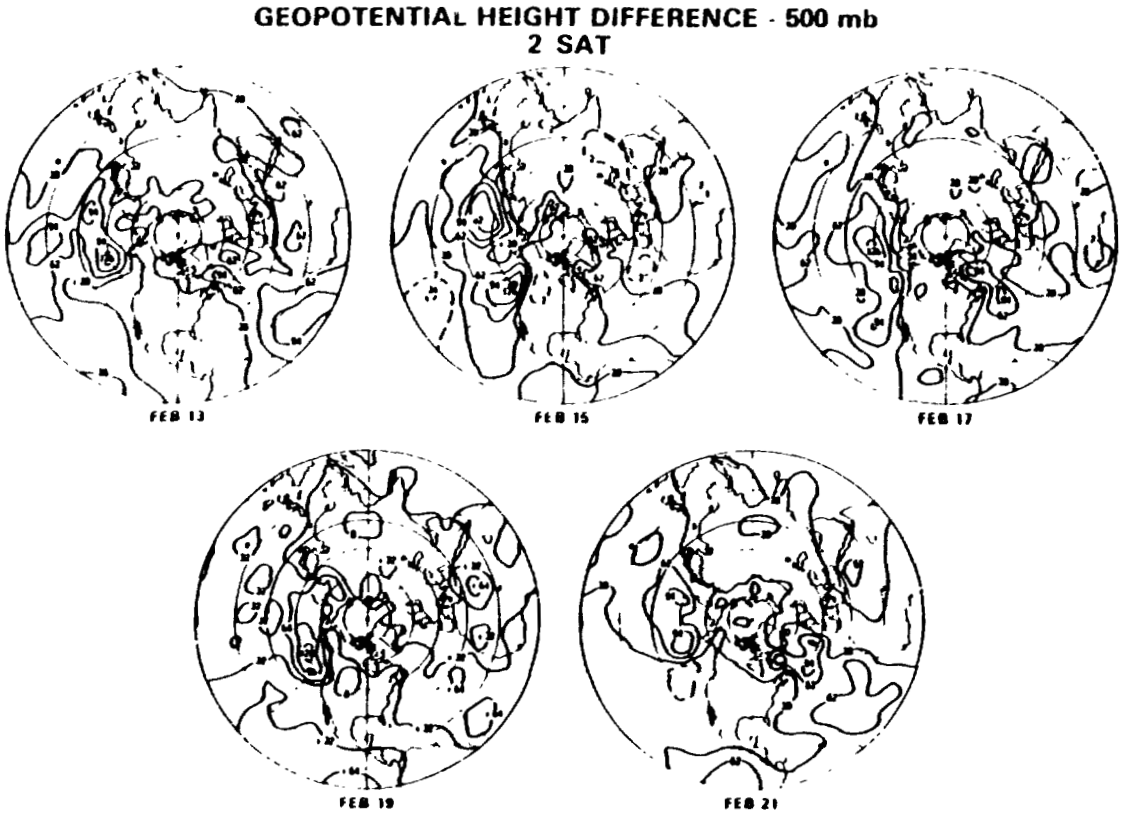


Fig 2 Initial state differences in 500 mb geopotential heights ϕ for the Northern Hemisphere between experiments 8566(2SAT) and 7578(NOSAT), $\phi_{33b}-\phi_{00}$. Positive differences are indicated by full contour lines, negative differences by dashed lines. The differences show a definite warm bias of satellite sounding temperatures, especially over the oceans.

2a) Differences on Feb. 1 through Feb. 11



2b) Differences on Feb. 13 through Feb. 21.

data from both satellites (called 2SAT) and the other withholding all satellite soundings (called NOSAT). The difference fields are given at the initial time of each one of the forecasts which were carried out for every experiment. A look at the fields indicates that large variations occur in the magnitude of initial state differences from one day to another. The magnitudes of the differences vary as a result of variations in data coverage and quality, as well as of the natural variability of the atmosphere. It is clear from the difference plots that day after day, differences of the order of 30-120 meters are produced in data-sparse regions of both the Pacific and Atlantic Oceans. Geopotential height differences of this magnitude correspond to 1.5° to 6°C temperature differences in the 1000 mb- to -500 mb thickness. The variations are of both signs, although there seems to be a systematic tendency for warmer temperatures in the SAT cycle over most of this assimilation test. Superimposing the satellite tracks (not shown) on the difference charts, we noticed that the largest differences occur mainly in the data-sparse areas along the tracks where sounding data have recently been inserted. It was also observed that in cases where the initial state differences are smaller, the forecast impact in terms of S_1 skill scores and RMS scores is generally also smaller. We intend to substantiate these qualitative observations by further quantitative studies.

From the nature of the initial state differences, one cannot say, a priori, whether the utilization of satellite data in a given case would produce beneficial impacts. It seems reasonable to assert at this point, however, that the magnitude of these differences is sufficiently large so that one would expect them to produce a certain number of significantly distinct forecasts.

Numerical Evaluation of Forecast Differences

The numerical measures of impact we used were S_1 skill scores and RMS differences. The meteorological fields we studied in particular were the sea-level pressure p_s and the height ϕ of the 500 mb geopotential surface. First the difference between values of the field produced by a model forecast and the field values of the NMC objective analysis at the same synoptic time were computed. These differences were computed for the synoptic time 48h and 72h after initial time.

The next step was to compute the RMS value of the difference. The actual impact measure consisted in subtracting the RMS difference for the experiment forecast from the RMS difference for the NOSAT forecast. A positive value of this difference then means positive impact of, or improvement due to satellite data for the given experiment and quantity, while a negative value means negative impact. The computations for skill score impacts were done in analogous fashion. The verification regions over which the impact was measured were North America and Europe.

Because of random influences on the results we discuss, these results had to be summarized in a way which would allow definite conclusions to be drawn. Such a summary is presented in Table II. For details we have again to refer to Ghil and Dilling (1977).

Table II. Summary of results for the numerical measures of impact.

Experiment No.	Code	Impact, % S1	Impact, % RMS	Statistical Significance (average/standard error) S1	Statistical Significance (average/standard error) RMS
7578	00	0	0		
8240	12	0.21	2.43	0.13	0.94
8405	22C	0.97	1.51	1.07	1.23
8352	22	1.83	6.31	1.01	1.82
8310	23	2.75	5.13	0.94	1.11
8447	22N	2.98	7.36	1.36	1.24
8574	31	1.20	3.66	0.66	1.19
8472	32	3.11	5.97	1.39	1.37
8545	33	4.74	9.57	1.94	2.26
8566	33a	4.46	12.41	1.85	2.77
8593	33b	4.79	10.02	2.05	2.49

The mean percentual impacts for each experiment will be denoted in the following discussion by x_j; the subscript j refers to the code number of the experiment. We shall denote the measure of statistical significance for the results by ζ_j. This measure is so defined (Ghil and Dilling, 1977) that we consider an experiment j to have yielded significant results, roughly speaking, if $\zeta_j \geq 1$; we have no confidence at all in the results if $\zeta_j \leq 0.5$, and high confidence if $\zeta_j \geq 2$.

Before discussing the results, it is important to remember that Experiment 8405 (code 22C) is a control experiment, as described in

Section 3, and that all the results have to be gauged against those of Experiment 8405. We discuss the results for skill scores first.

In Table II we notice that the measure of statistical significance for Experiment 8405 is $\zeta_{22C} = 1.07$. Hence the results of the DIM experiment 8240, with $\zeta_{12} = 0.13$, have no confidence whatever attached to them (also $\zeta_{12} < 0.5$), while those of the SCM experiments 8352, 8310, and 8447, with $\zeta_{22} = 1.01$, $\zeta_{23} = 0.94$, and $\zeta_{22N} = 1.36$ respectively, can be considered as marginally significant. The corresponding mean percentual impacts from Table II are $x_{22C} = 0.97$ for the control experiment 8405, and $x_{12} = 0.21$ for the DIM experiment 8240, which is practically negligible; they are $x_{22} = 1.83$, $x_{23} = 2.75$, and $x_{22N} = 2.98$ for the SCM experiments 8352, 8310, and 8447 respectively, with $x_{22N} \simeq x_{23} \simeq 1.5x_{22}$, $x_{22} \simeq 2x_{22C}$. We already notice a strong correlation between measure of impact and measure of confidence. Moreover, the control experiment 8405 shows that the DIM experiment totally failed to produce favorable impacts, while the SCM experiments produced results considerably better than the performance level set by the control experiment, in fact twice as good, at least. Also Experiment 8310, which used a large amount of data, obtained from two satellites, produced results which were 50 percent better in mean percentual impact than those of Experiment 8352, utilizing the same assimilation method (SCM) but data from Nimbus-6 only. Furthermore, Experiment 8447, in which no satellite data were inserted over land, reflecting higher reliance on conventional data there, produced results comparable to and even better than those of Experiment 8310, while using only data from one satellite (Nimbus-6).

We discuss now the SAM experiments, 8574 through 8593. The comparison of skill score results, in both impact and significance, of Experiment 8574, which used only NOAA-4 data, Experiment 8472 which used the more abundant Nimbus-6 data, and the remaining SAM experiments, 8545, 8566, 8593, which used data from both satellites, confirms the importance of data quantity, independently of the assimilation method. We have in fact $\zeta_{31} = 0.66 < \zeta_{32} = 1.39 < \zeta_j$, $j = 33$, 33a, 33b, with $\zeta_j \simeq 2$; similarly $x_{31} = 1.20 < x_{32} = 3.11 < x_j \simeq 4.65$. Thus Experiment 8472 is comparable to the best SCM experiments, 8310 and 8447, while Experiment 8574 produces somewhat poorer results.

Results of the three SAM experiments which utilized fully the available data, 8545, 8566, and 8593, are remarkably similar. We have $\zeta_{33} = 1.94$, $\zeta_{33a} = 1.85$, and $\zeta_{33b} = 2.05$; these give us all high statistical confidence in the mean impacts, being very close to and even larger than 2. The mean impacts are $x_{33} = 4.74$, $x_{33a} = 4.46$,

and $x_{33b} = 4.79$, that is close to 5 percent. This is certainly not a very large impact, but is quite comparable to improvements in numerical weather prediction which have been considered as important over the last decade; it corresponds approximately to the ability of making a 60h forecast of accuracy which equals that of today's operational 48h forecast.

We remark in passing that attempts at removing bias, (cf. Ghil and Dilling, 1977) of satellite-derived temperatures, either at observation points (8566), or at the correction point (8593), did not make much difference in the results. The use of SAM itself, however, certainly did make a difference. It is very interesting to consider the SCM experiment which is similar in other respects to the three SAM experiments we are in the process of discussing, namely Ecperiment 8310; we immediately notice that the two quantities measuring result, ζ_{23} and x_{23}, have a value which is almost exactly half the corresponding representative value for SAM, $\zeta_{23} = 0.94$, versus 2, and $x_{23} = 2.75$, versus 5.

Values of ζ_{12} and x_{12} for DIM experiment 8240 can be considered zero for all practical intents and purposes of this discussion. We observe at this point that, in a certain sense, SCM is a low-order approximation to SAM, in which the matrix $\underset{\sim}{A}$ of Section 2 is approximated by a diagonal matrix (compare Rutherford, 1972).

Results for RMS errors strongly support those presented here for skill scores; they are given in the fourth and sixth columns of Table II. The statistical significance of the RMS results is influenced by data quantity and assimilation method in the same way as the S_1 results, and so are the mean impacts; the values of RMS mean impacts are of the order of 10 to 12 percent for the best SAM experiments, and 5 to 7 percent for the SCM experiments.

It appears in general that differences in initial states correlate reasonably well with impact on forecasts from these initial states (v. Subsection 4a). Our numerical measures of impact also seem to correlate positively with improvements in the capability of predicting local weather when using the large-scale numerically predicted fields for guidance (v. Subsection 4d).

Synoptic Evaluation of Forecast Differences

It is much more difficult to carry out a subjective comparison of prognostic fields based on synoptic criteria than to compute a numerical measure of the difference between such fields. Therefore we have to limit ourselves at this point to the presentation of a comparison of the results for experiment 8566 (code 33a) with those for the NOSAT experiment 7578. In this and the next subsection we shall again refer to experiment 8566 simply as 2SAT.

The synoptic evaluation of forecasting impact for the satellite sounding data was performed by directly comparing the prognostic charts generated from 2SAT and NOSAT initial conditions with each other and with the corresponding NMC analyses for verification. Comparisons were made every 12 hours for a variety of prognostic fields.

The results of our comparisons for the sea-level pressure and 500 mb height fields verified over North America at the end of each 72-hour forecast are presented in Table III. The difference plots on which these results are based are given in Atlas and Sakal (1977). The table was prepared by first noting the regions of moderate to large differences between the 2SAT and NOSAT prognostic fields and then comparing the actual errors of each system in these regions.

Table III. Summary of regional 72h forecast synoptic impacts over North America. Only impacts of 8mb or larger in sea-level pressure and of 96 meters or larger in 500mb geopotential have been included.

A. Sea-Level Pressure

Forecast from	Beneficial Impacts (in mb)	Negative Impacts (in mb)
Feb. 1	+12	-12
3	+ 8	
3	+16	
5		- 8
7	+16	-16
9	+16	
9	+ 8	
11	+12	
13	+ 8	
19	+ 8	
21	+ 8	

B. 500mb Geopotential

Forecast from	Beneficial Impacts (in meters)	Negative Impacts (in meters)
Feb. 1		- 96
3	+160	
5	+ 96	
7		-160
9	+128	
11	+ 96	
13	+ 96	
19	+128	
21	+ 96	

The table shows that moderate to large forecasting impacts occur in nine of the eleven cases and that more than two thirds of these impacts are beneficial. As a result, there are 40-75% reductions of specific forecasting errors at 500mb and 40-100% reductions of specific errors in the sea-level pressure forecast.

One example of a beneficial forecasting impact over North America is presented in Figure 3. Figures 3a and 3b depict the 72-hour 500mb prognostic charts from February 19 for the NOSAT and the 2SAT case respectively. The corresponding NMC analysis is depicted in Figure 3c. A comparison of these three charts reveals that the NOSAT prognosis is considerably slow in its movement of a trough into the Midwest of the United States and of a ridge off the East Coast. Both of these systems are displaced further east, in better agreement with the analysis, by the 2SAT prognosis. A detailed synoptic study of the 72-hour prognostic fields over North America was also carried out for the forecast started on 03Z Feb. 9 (Atlas and Sakal, 1977). It showed considerable improvement in the steering of weather systems by the 2SAT forecast.

Differences in Local Precipitation Forecasts

The ultimate objective in evaluation of satellite data impact is to determine what the impact means in terms of the accuracy of actual weather predictions that might be issued by a local weather forecaster. In order to assess this impact, we developed a computerized forecasting model (CFM) which simulates a human forecaster's interpretation of a set of prognostic charts under the restrictive condition that no modification to the numerical prognostic output is to be made. The CFM was found quite satisfactory when applied to perfect prognostic fields, i.e., to analyzed observed fields, and

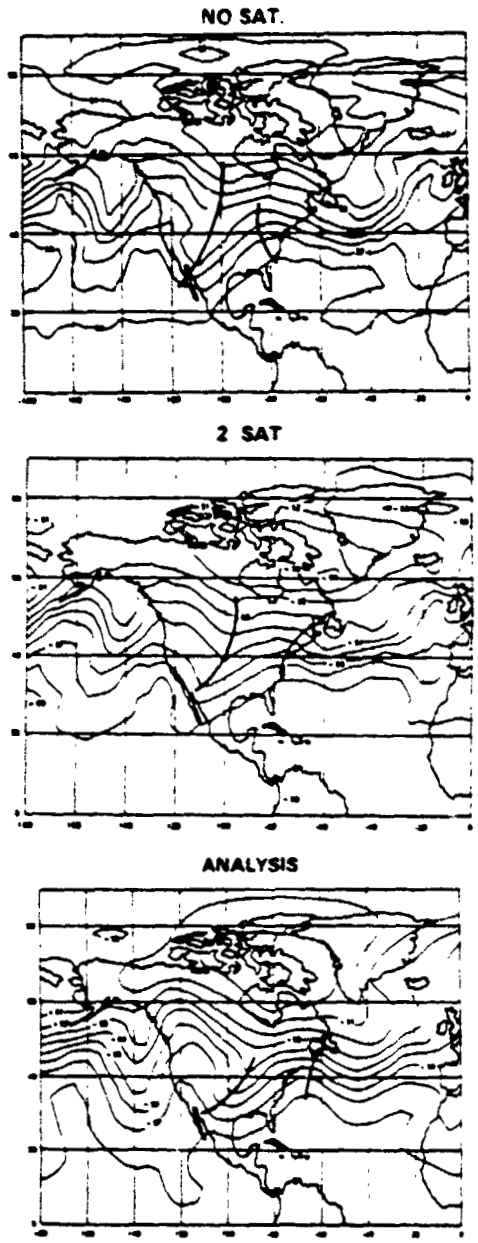


Figure 3. Synoptic differences between numerical predictions from an initial state containing temperature sounding information and from one not containing such information. The case shown is the 72h forecast from 03Z February 19 for the 2SAT experiment 8566 (Fig. 3b) and the NOSAT experiment 7578 (Fig. 3a). The fields shown are the 500 mb geopotential height fields over North America. For comparison, Fig. 3c gives the operational analysis of NMC at 00Z February 22. Clearly the 2SAT forecast is closer to the analysis than the NOSAT forecast.

verified against observed precipitation; it also agreed quite well with predictions made by qualified human forecasters using the same model prognostic fields. The CFM is described and further details are given in Atlas and Sakal (1977).

CFM local precipitation forecasts were generated for the 11 test cases from both the 2SAT and the NOSAT prognostic fields; they were generated at 24h, 48h and 72h after initial time for 128 cities uniformly distributed over the United States. Thus a total of $2 \times 3 \times 11 \times 128 = 8,448$ local precipitation forecasts was made for the comparison we discuss.

Comparisons of precipitation forecasts for these 128 cities are presented in Table IV. The number of correct precipitation forecasts for the 2SAT and NOSAT systems are shown only for those cases in which the two systems gave rise to different forecasts. The systems actually differed for about 10 percent of all forecasts and the results given cover the cases when one system forecasts a precipitation event while the other one forecasts no precipitation. The number of different forecasts for each day is shown for all the 128 cities considered together in the lower horizontal row. At 24 and 48 hours there is a 12 percent improvement in the 2SAT system, while at 72 hours the forecast improvement was 19 percent; the percentual improvement is computed with respect to the total number of different forecasts. We examine in the upper row a separate list of 17 cities in the Midwest, selected because they are less influenced by coastal effects, mountains, or the warmer Gulf convective systems, which the CFM has difficulty accounting for in its present form. In the case of these cities we see an improvement of better than two to one in forecasting for 24 and 72 hours and an even larger improvement for 48 hours.

Table IV. Summary of results for local precipitation forecasting in cases where use of 2SAT and NOSAT prognostic fields produced different forecasts. "2SAT better" means that the CFM produced a correct precipitation forecast from 2SAT prognostic fields, and an incorrect one from NOSAT fields; "NOSAT better" means the opposite was the case.

	24 h		48 h		72 h	
	2SAT Better	NOSAT Better	2SAT Better	NOSAT Better	2SAT Better	NOSAT Better
Midwest (17)	11	4	11	1	7	3
All US cities (128)	82	64	68	53	85	58

To exhibit the correlation between numerical impact measures and impact on local weather forecasts we present Tables V and VI. In Table V we give S_1 skill score impacts verified at 72 hours over North America for each forecast separately. In Table VI we give the corresponding precipitation forecast impacts. We observe large numerical impacts in both ϕ and p_s for the forecasts from February 3, 9, 11 and 13; they are all positive, and large when compared to those of the control experiment 8405. The precipitation forecast impacts at 72 hours are large when compared to the NOSAT experiment for February 9 and 11, and moderate, but still positive for February 3 and 13. A large positive impact also occurs on February 19; while the numerical measures did not show a strong impact of sounding data for this forecast, the synoptic comparison (Subsection 4c) did show a significant improvement in the 2SAT prognostic fields. A large, negative impact on precipitation forecast at 72 hours occurs for February 15, when the numerical measures show little impact, if at all.

Table V. Daily skill score impacts verified over North America at 72 hours.

	500 mb Heights		Sea-Level Pressure	
Date	2SAT (8566)	Control Experiment (8405)	2SAT (8566)	Control Experiment (8405)
Feb. 1	1.8	-3.5	-4.9	-12.0
3	9.4	3.4	8.4	4.7
5	-2.2	0.8	-0.3	-2.8
7	0.7	-1.1	10.2	0.3
9	7.8	1.2	13.5	1.4
11	5.9	-2.1	16.4	-1.2
13	6.8	1.1	10.5	5.2
15	1.9	-3.3	1.5	-3.7
17	-4.5	-1.8	-5.7	-4.5
19	4.7	3.9	-2.8	5.8
21	5.4	-0.9	5.2	-4.3
	3.43±1.3	-0.20±0.8	4.73±2.3	-1.01±1.6

From these comparisons we conclude to the need for much more study of the interrelationship between initial state differences and different forecast impact measures. The preliminary evidence, although incomplete, seems to point in the direction of positive correlation between large impacts on initial state determination in data-sparse areas, on the one hand, and large impacts on the different quantities measuring the accuracy of forecasts at 48 and 72 hours over data-rich areas, on the other.

Table VI. Daily results of local precipitation forecasting for cases of different forecasts. The results refer to 128 U.S. cities and are verified at 72 hours.

Date	2SAT Better	NOSAT Better	
Feb. 1	3	6	
3	8	7	
5	7	5	
7	7	4	
9	17	8	
11	3	0	
13	5	3	
15	8	16	
17	5	3	
19	21	3	
21	1	3	
	85	58	Totals

CONCLUDING REMARKS

We carried out a study to assess the effects of temperature sounding data from satellite-borne instruments on short-range (1- to 3-day) weather forecasting. The instruments providing the data were operational and experimental temperature sounders aboard the NOAA-4 and Nimbus-6 satellites.

We performed a number of experiments in which satellite data were assimilated to produce initial states for numerical forecasts by a number of different methods. The experiments used data from the DST-1 period. The methods considered were direct insertion (DIM), asynoptic successive corrections (SCM) and time-continuous statistical assimilation (SAM).

Results of our experiments allow us to draw a number of conclusions: (i) satellite-derived temperature data can have a modest, but statistically significant positive impact on numerical weather forecasts, as verified over the continents of the Northern Hemisphere; (ii) this impact is highly sensitive to the quantity of the data—the impact of a two-satellite system is larger than that of one satellite by an amount roughly proportional to the quantity of data provided; (iii) the assimilation method plays a major role in the magnitude of the impact for the same data—direct insertion had a practically null impact, while SCM provided about half the impact obtained with SAM.

The impact for the best method tested and the full amount of data available was about 5 percent in S_1 skill score and 12 percent

in RMS errors; these correspond to the possibility of an extension of about 8-12h in the usefulness of numerical weather prediction (NWP) in the range between 48h and 72h. There are indications that local weather forecasts using large-scale NWP results as guidance can be similarly improved.

We are presently further refining our statistical assimilation method of the sounding data and of applying it to data from the entire DST-6 period. We intend also to adapt SAM to the assimilation of cloud-track wind data from geostationary satellites, and we expect eventually to apply it to FGGE (First Global GARP Experiment) data.

Our results point to two major areas in which improvements can bring about larger impacts of satellite-borne sounding systems. First, instrument development to improve the accuracy and vertical resolution of the sounder temperature profiles themselves. At present, vertical temperature profiles derived from satellite sounding radiance data have 2-2.5°C RMS errors when compared with co-located radiosonde profiles. Although this accuracy falls short of meeting GARP (Global Atmospheric Research Program) requirements, we have shown that the data are still capable of producing modest impacts when properly utilized. However, the deficiencies in the quality of the data can only partially be compensated for by special retrieval and assimilation methods. There is certainly a need for continuous data monitoring and for continued development of retrieval and of assimilation methods. Moreover, close interaction between retrieval and assimilation methods seems to hold great promise. But the most important requirement is still that of instrument packages with greater accuracy and resolution, capable of providing reliable soundings under all atmospheric conditions.

Second, numerical prediction models themselves must be improved to make better use of the sounding data. In order to successfully assimilate asynoptic data and have them contribute to more accurate forecasts, it is necessary that the model be able to convey information accurately over extended distances and periods of time from one region of the globe to another. Improvement in forecasting models can only proceed by a judicious combination of higher grid resolutions, more accurate numerical discretization methods, and better representations of atmospheric processes in the model.

We hope that a concerted effort in designing better observing instruments and systems, refining the methods for processing and assimilating their observations, and developing better numerical models

will lead to considerable improvements in numerical weather prediction and to better understanding of the atmospheric circulation; these are the goals of the Global Atmospheric Research Program.

ACKNOWLEDGEMENTS

It is a pleasure to acknowledge all the contributors to the GISS Sounding Temperature Impact Test 1976.

REFERENCES

Atkins, M.J., and M.V. Jones (1975). An experiment to determine the value of satellite infrared spectrometer (SIRS) data in numerical forecasting. Meteor. Mag., 104, 125-142.

Atlas, R., and D. Sakal (1977). Evaluation and verification tests. Chapter 5 of Halem et al. (1977), pp. 5.1 - 5.81.

Bengtsson, L. (1975). Four-Dimensional Data Assimilation of Meteorological Observations. GARP Publ. Ser., No. 15, WMO/ICSU, Geneva, Switzerland, 76 pp.

_____, and N. Gustavsson (1971). An experiment in the assimilation of data in dynamical analysis. Tellus, 23, 328-336.

_____ (1972). Assimilation of nonsynoptic observations. Tellus, 24, 383-399.

Charney, J., M. Halem and R. Jastrow (1969). Use of incomplete historical data to infer the present state of the atmosphere. J. Atmos. Sci., 26, 1160-1163.

Cressman, G.P. (1959). An operational objective analysis system. Mon. Wea. Rev., 85, 367-374.

Eliassen, A. (1954). Provisional report on calculation of spatial covariance and autocorrelation of the pressure field. Report No. 5, Inst. Weather and Climate Res., Acad. Sci., Oslo, 11 pp.

Gandin, L.S. (1963). Objective Analysis of Meteorological Fields. Gidrometeorologischeskoe Izdatel'stvo (GIMIZ), Leningrad. English translation by Israel Program for Scientific Translations, Jerusalem 1965 [available from NTIS], 242 pp.

Ghil, M. (1975). Initialization by compatible balancing. Report 75-16, Inst. Comp. Appl. Sci. Eng., NASA Langley Research Center, Hampton, VA 23665, 38 pp.

_____, and R. Dilling (1977). Analysis and assimilation. Chapter 3 of Halem et al. (1977); pp. 3.1 - 3.150.

_____, B. Shkoller and V. Yangarber (1977a). A balanced diagnostic system compatible with a barotropic prognostic model. Mon. Wea. Rev. 105, 1223-1238.

_____, R. Dilling and H. Carus (1977b). A statistical method for the time-continuous assimilation of satellite-derived temperatures, Proc. 5th Conf. Probability Statistics Atmos. Sci., Amer. Meteor.

Soc., Boston, 320-324.

Halem, M., M. Ghil, R. Atlas, J. Susskind and W.J. Quirk, eds., (1977). The GISS Sounding Temperature Impact Test. NASA Institute for Space Studies, Goddard Space Flight Center, New York, NY 10025, 328 + xiv pp.

Hayden, C.M. (1973). Experiments in the four-dimensional assimilation of Nimbus 4 SIRS data. J. Appl. Meteor., 12, 425-436.

Murphy, A., and D. Williamson, Coordinators (1976): Weather Forecasting and Weather Forecasts: Models, Systems and Users. Notes from a Colloquium held at the National Center for Atmospheric Research, Summer 1976, NCAR/CQ-5+1976-ASP, volumes I and II, 900 pp.

Rutherford, I.D. (1972). Data assimilation by statistical interpolation of forecast error fields, J. Atmos. Sci., 29, 809-815.

_____ (1973). Experiments in the updating of P.E. forecasts with real wind and geopotential data. Proc. 3rd Conf. Probability Statistics Atmos. Sci., Amer. Meteor. Soc., Boston, 198-201.

Smagorinski, J., K. Miyakoda and R.F. Strickler (1970). The relative importance of variables in initial conditions for dynamical weather prediction. Tellus, 22, 141-157.

Smith, W.L., and H.M. Woolf (1976). The use of eigenvectors of statistical covariance matrices for interpreting satellite sounding radiometer observations. J. Atmos. Sci., 33, 1127-1140.

Somerville, R.C.J., P.H. Stone, M. Halem, J.E. Hansen, J.S. Hogan, L.M. Druyan, G. Russell, A.S. Lacis, W.J. Quirk and J. Tenenbaum (1974). The GISS model of the global atmosphere. J. Atmos. Sci., 31, 84-117.

Susskind, J., and D. Edelmann (1977). Sounding temperature studies. Chapter 2 of Halem et al. (1977), pp. 2.1 - 2.82.

Werbowetzki, A. (1975). Indirect sounding of the atmosphere from NOAA spacecraft - Regression after categorization, method and results. Proc. 4th Conf. Probability Statistics Atmos. Sci., Amer. Meteor. Soc., Boston, 165-170.

Paper 14

NASA Technical Memorandum 80608 December 1979

October 1979

Band Fourth Order Model Preliminary Documentation

Eugenia Kalnay-Rivas; David Hoitsma

Abstract

A preliminary documentation of the Band Fourth Order Global Atmospheric Model has been compiled. It contains the dynamic equations for atmospheric motion, finite difference equations, a discussion of the structure and flow chart of the program code, a copy of the code and three relevant papers.

Paper 15

Reprinted from JOURNAL OF THE ATMOSPHERIC SCIENCES, Vol. 36, No. 7, July 1979
American Meteorological Society
Printed in U. S. A.

Numerical Simulation of Ozone Production, Transport and Distribution with a Global Atmospheric General Circulation Model

MICHAEL E. SCHLESINGER[1] AND YALE MINTZ[2]

Department of Atmospheric Sciences, University of California, Los Angeles 90024

(Manuscript received 27 October 1978, in final form 5 March 1979)

ABSTRACT

The production, transport and distribution of ozone are simulated for a January with a global atmospheric general circulation model. In this model the ozone influences the radiational heating as well as the photochemical ozone production and destruction, the radiational heating influences the atmospheric circulation, and the circulation redistributes the ozone.

The model has fairly successfully simulated the synoptic and time-averaged observed large-scale fields of temperature, mass, and velocity in the troposphere and stratosphere, although there are some deficiencies. In particular, the simulated temperatures are too cold in the lower and middle stratosphere in the polar regions, the sea level pressure is too high in the Arctic and in the Antarctic circumpolar trough, and the flow field in the middle-latitude troposphere does not show the observed wavenumber 3.

Despite these shortcomings, the model has simulated the observed high correlation of synoptic and time-averaged total ozone with the tropospheric height field in middle latitudes, with the ozone maxima and minima, respectively, located at the troughs and ridges of the tropospheric waves. The deficiencies which are seen in the time-averaged O_3 distribution are attributable to recognized deficiencies of the general circulation model.

In the tropics there is a vertically integrated transport of O_3 from the summer to the winter hemisphere which is almost entirely produced by the mean-meridional circulation. In the middle latitudes, in both hemispheres, O_3 is transported toward the equator by the mean-meridional circulation and toward the poles by the zonal eddies; but the eddy transport dominates, so that the net O_3 transport is poleward. In the high latitudes in both hemispheres, there is a reversal in the directions of the two components of the O_3 transport; but here the transport by the mean-meridional circulation dominates, so that the net transport continues to be poleward.

In the individual latitudes, the zonally integrated vertical transport of ozone is dominated by the transport by the mean-meridional circulation; but integrated over the globe, the vertical O_3 transport is dominated by the eddy transport. Between 20 and 31 km elevation, the globally integrated vertical O_3 transport is a countergradient transport with respect to the globally integrated O_3 mixing ratio.

The divergence of the O_3 transport maintains the ozone below its photochemical equilibrium concentration in the tropics and subtropics, and the convergence of the O_3 transport maintains the ozone above its photochemical equilibrium concentration in the middle and high latitudes of both hemispheres. In this way, both the atmospheric motions and the O_3 photochemistry determine the O_3 sources and sinks.

The globally integrated photochemical production of ozone exhibits variations with periods of a day and less. These high-frequency oscillations are due to the quasi-stationary longitudinal variation in the ozone that is produced by the O_3 transports.

1. Introduction

Photochemical theory explains the existence of ozone in the earth's atmosphere, but not its observed distribution in space and time. According to the photochemical theory developed by Chapman (1930a,b,c), solar ultraviolet (UV) radiation photodissociates molecular oxygen, and the resulting atomic oxygen combines with molecular oxygen to produce ozone. The ozone is also photodissociated by solar UV radiation and in addition, is destroyed by recombination with atomic oxygen (the Chapman mechanism). Recent findings have shown that O_3 destruction by the Chapman mechanism is responsible for only about 20% of the total destruction which is required to balance the global O_3 production (Johnston and Whitten, 1973). Several possible chemical destruction mechanisms have been considered, but it now appears that the oxides of nitrogen and hydrogen are the predominant O_3 destroying agents.

[1] Present affiliation: Department of Atmospheric Sciences and Climatic Research Institute, Oregon State University, Corvallis 97331.

[2] Present affiliation: Laboratory for Atmospheric Sciences, Code 911, NASA Goddard Space Flight Center, Greenbelt, MD 20771

0022-4928/79/071325-36$13.00

Photochemical theory predicts a decrease of total O_3 from the equator toward higher latitudes and a summer maximum and winter minimum of total O_3. However, observations show that the total O_3 increases from the equator toward higher latitudes and has a winter–spring maximum and summer–fall minimum (Dütsch, 1969). It has been proposed that this would be the result of a net poleward transport of ozone by the mean-meridional circulation if the air moves poleward at high altitudes, where the O_3 mixing ratio is large, and equatorward at low altitudes, where the O_3 mixing ratio is small (Dobson *et al.*, 1929; Brewer, 1949; Dobson, 1956, 1973). It has also been proposed that zonal eddies in the atmospheric motion field would produce a net poleward O_3 transport if, in the stratosphere, the poleward moving and descending air parcels contain more O_3 than the equatorward moving and ascending air parcels (Newell 1961, 1963a,b, 1964a,b; Newell *et al.*, 1966, 1973).

To determine the O_3 transports, one must know the O_3 concentration and the horizontal and vertical velocity components as functions of space and time. At present, our knowledge of these is not sufficiently accurate to calculate the O_3 transports from the observations.

A number of simulation studies have been made of O_3 production, transport and distribution with two-dimensional (i.e., zonally averaged) atmospheric models. A recent study of this kind is the one by Harwood and Pyle (1977). But these models have two shortcomings. The first is that the horizontal and vertical transports of momentum, heat and ozone by large-scale atmospheric eddies (which play a critical role in determining the sources, sinks and distributions of these variables) must be parameterized as functions of the zonally averaged variables; but there is no general agreement as to how this can be correctly done. The second shortcoming is that these models tell us nothing about the zonal variations of the O_3 distribution nor about the processes which produce those east-west variations.

We can avoid these shortcomings by simulating the O_3 production, transport and distribution with a three-dimensional general circulation model in which the large-scale atmospheric eddies are explicitly resolved. Studies of this kind have been made by Hunt and Manabe (1968), Hunt (1969), Clark (1970), Mahlman (1972), London and Park (1973, 1974), Newson (1974) and Cunnold *et al.* (1975). In all of those studies, however, some of the interactive processes which determine the O_3 distribution were either ignored or were highly constrained. Hunt and Manabe (1968), Mahlman (1972), London and Park (1973) and Newson (1974) calculated the three-dimensional O_3 transports as a function of the simulated atmospheric motions, but ignored the photochemical production and destruction of ozone and the influence of the predicted ozone on the radiational heating and, consequently, its feedback on the atmospheric motions. Hunt (1969) and London and Park (1974) included the O_3 photochemistry as well as the transports, but did not use the predicted ozone when calculating the radiational heating and the atmospheric motions. Clark (1970) and Cunnold *et al.* (1975) included the transport, the photochemistry and the radiational heating influence of the predicted ozone on the stratospheric circulation; but in their general circulation models, the tropospheric circulation, to which the stratospheric circulation is strongly coupled, was highly constrained by an empirically based Newtonian-type heating of the troposphere.

In the present study we have sought to simulate the global O_3 production, transport and distribution with a numerical general circulation model in which the O_3 photochemistry, the radiational heating and the O_3 transports are coupled and interactive, and in which the tropospheric heating is realistically modeled. We have used a version of the atmospheric general circulation model that was developed by the UCLA general circulation research group for this and other purposes, and in which ozone is one of the prognostic variables. In this model, the ozone influences not only its photochemical production and destruction, but also the atmospheric radiational heating; this radiational heating influences the atmospheric circulation, and the circulation redistributes the ozone.

2. Description of the model

A detailed account of the design of the atmospheric general circulation model has been given by Arakawa and Mintz (1974) and Arakawa and Lamb (1977). What follows is a brief description of the model.

a. Prognostic and diagnostic variables

The principal prognostic variables of the model are the *horizontal velocity*, *temperature* and *surface pressure*, governed, respectively, by the horizontal momentum equation, the thermodynamic energy equation and the surface pressure tendency equation. With the mass continuity equation and the hydrostatic equation, and appropriate boundary conditions, these equations form a closed system for an adiabatic and frictionless atmosphere.

But the general circulation of the atmosphere is the large-scale, thermally driven field of motion in which there are interactions between the heating field and the motion field. Therefore, several additional prognostic variables, with corresponding governing equations and appropriate boundary

conditions, must be added to simulate the heating. Because the radiational heating of the stratosphere is strongly governed by ozone, which varies in space and time, *ozone* was made a prognostic variable of the model, governed by an O_3 continuity equation with parameterized sources and sinks. An even more important prognostic variable which influences the heating is *water vapor*, which is governed by the water vapor continuity equation. Because the source and sink terms in the water vapor continuity equation and in the thermodynamic energy equation depend on subgrid-scale cumulus convection, the model has a parameterized penetrative cumulus convection, based on the theory of Arakawa and Schubert (1974), which interacts with a parameterized planetary boundary layer. The *boundary-layer depth* and the *temperature discontinuity* and *moisture discontinuity* at the top of the boundary layer are also prognostic variables of the model (see Randall, 1976).

Because there can be a large diurnal heat storage in the soil, because water stored in the soil can be a significant source of water vapor for the atmosphere, and because snow lying on the ground can have a large influence on the surface albedo and also be a significant water vapor source, the *soil temperature*, *soil water storage* and *mass of snow on the ground* are prognostic variables of the model, governed by energy, water and snow budget equations for the ground.

In addition to the prognostic variables, the model has many diagnostic variables (physical state parameters whose magnitudes, rather than whose time rates of change of magnitude, are determined by the governing equations). Among the more important of these are the different types of clouds which form with the release of latent heat. These clouds affect the solar and longwave radiational heating of the air. Any model layer below 100 mb which is saturated is taken to be filled with cloud. In addition, if the top of the planetary boundary layer is saturated it has a sublayer of stratus cloud. Furthermore, if the parameterized cumulus convection has penetrative cumulonimbi which detrain at levels higher than 400 mb, the layer in which the detrainment takes place is taken to be filled with cloud. Clouds colder than −40°C or above 400 mb are taken to be ice crystal clouds, and clouds warmer than −40°C and below 400 mb are taken to be water droplet clouds.

b. Physical processes

In addition to the three-dimensional transports of the prognostic variables, the sources and sinks of the variables are calculated.

For the horizontal velocity, the dominating source and sink terms in the horizontal momentum equation are a frictional drag at the surface of the earth, with a subgrid-scale vertical eddy flux of horizontal momentum within the planetary boundary layer; and a subgrid-scale vertical eddy flux of horizontal momentum within the free atmosphere which depends on the parameterized cumulus convection.

For the temperature, the source and sink terms in the thermodynamic energy equation are a solar and longwave radiational heating; a sensible heat flux across the earth's surface, with a subgrid-scale vertical eddy flux of heat and a release of heat of condensation within the planetary boundary layer; a subgrid-scale vertical flux of sensible heat and a release of heat of condensation by the parameterized cumulus convection; and a heating by large-scale condensation and cooling by evaporation of the falling raindrops. The solar and longwave radiation parameterizations are modified and extended versions of the methods developed by Katayama (1972) for the UCLA tropospheric general circulation models. For the solar radiation, the effects of water vapor and ozone absorption, molecular scattering, and the scattering and absorption by clouds are included. For the longwave radiation, up to the 30 km level, the emission and absorption by water vapor, carbon dioxide and ozone are included; and the clouds are treated as blackbodies if they are water droplet clouds and as grey bodies, with emissivity of one-half, if they are ice crystal clouds. But above 30 km, the longwave radiational cooling is calculated using the approximation of Dickinson (1973). Schlesinger (1976a, Appendix A) has given a detailed description of the entire radiation calculation scheme in this general circulation model.

For the water vapor, the source and sink terms in the water vapor continuity equation are an evapotranspiration at the earth's surface, with a subgrid-scale vertical eddy flux of water vapor within the planetary boundary layer; a subgrid-scale vertical eddy flux of water vapor and condensation by the parameterized cumulus convection; and large-scale condensation and evaporation of the falling raindrops.

For the ozone, the source and sink terms in the O_3 continuity equation are a highly simplified photochemistry above the 100 mb level, following Cunnold *et al.* (1975), in which O_3 is the only prognostic variable; and a subgrid-scale vertical eddy flux of O_3 within the planetary boundary layer with destruction by heterogeneous chemical reactions at the earth's surface parameterized following Cunnold *et al.* (1975).

The photochemical production and destruction of ozone is based on the Chapman reactions

$$O_2 + h\nu \xrightarrow{j_1} 2O,$$

$$O + O_2 + M \xrightarrow{k_1} O_3 + M,$$

$$O_3 + h\nu \xrightarrow{j_2} O_2 + O,$$

$$O + O_3 \xrightarrow{k_2} 2O_2,$$

and the $NO-NO_2$ catalytic cycle

$$NO + O_3 \xrightarrow{k_3} NO_2 + O_2,$$

$$NO_2 + O \xrightarrow{k_4} NO + O_2,$$

$$NO_2 + h\nu \xrightarrow{j_3} NO + O.$$

The HO_x reactions of the Cunnold *et al.* (1975) photochemical model are not included here because their results showed that these reactions are important only above 45 km, which would involve only our model's uppermost layer.

The chemical reaction rates k_i are

$$k_1 = 6.6 \times 10^{-35} \exp(510/T)[\text{cm}^6\ \text{s}^{-1}],$$

$$k_2 = 1.9 \times 10^{-11} \exp(-2300/T)[\text{cm}^3\ \text{s}^{-1}],$$

$$k_3 = 9 \times 10^{-13} \exp(-1200/T)[\text{cm}^3\ \text{s}^{-1}],$$

$$k_4 = 9.1 \times 10^{-12}[\text{cm}^3\ \text{s}^{-1}],$$

and the photodissociation rates j_i are

$$j_i(z) = \int_0^\infty \alpha_{\lambda,i} I_{\lambda,\infty} \exp\left[-\sec\zeta \sum_k \alpha_{\lambda,k} \int_z^\infty n_k(z)dz\right] d\lambda.$$

Here $I_{\lambda,\infty}$ is the extraterrestrial monochromatic photon flux per unit wavelength, $\alpha_{\lambda,i}$ the absorption cross section of species i at wavelength λ, $n_k(z)$ the number density of species k at altitude z, ζ the solar zenith angle, and the summation extends over all species k that absorb in the spectral region where species i photodissociates. The photodissociation rates are computed by a linear bivariate interpolation of precomputed values of j_i as a function of the O_2 and O_3 absorber amounts (Schlesinger, 1976a).

Because the three-dimensional distribution of NO_2 is not known, the NO_2 distribution is prescribed to vary only in the vertical following the one-dimensional profile calculated by McElroy *et al.* (1974).

The transport of ozone by the horizontal velocity component is calculated with a space-centered scheme. For a downgradient transport, an arithmetic mean of the O_3 mixing ratios is used; for a countergradient transport, a harmonic mean is used. This scheme reduces the spurious generation of negative O_3 mixing ratios produced by time- and space-truncation errors, and any remaining generation of negative O_3 mixing ratios is prevented by decreasing the magnitude of the horizontal O_3 flux between neighboring grid points. The scheme has a generally diffusive character because, for given O_3 mixing ratios and mass transport, the O_3 transport is greater with the arithmetic mean than with the harmonic mean.

The transport of O_3 by the vertical velocity was initially calculated with the above scheme; but, for the reason given below, this scheme was replaced by a space-centered scheme which is based on the conservation of the global integral with respect to mass of the logarithm of the O_3 mixing ratio (Schlesinger, 1976a).

c. Prescribed Parameters

The following are the prescribed parameters of the model:

- Radius, surface gravity and rotation speed of the planet
- Solar constant and orbital parameters of the planet
- Total atmospheric mass
- Thermodynamic and radiation constants of the atmospheric gases and clouds
- Geographic distribution of open ocean, ice-covered ocean, ice-free land and land covered by glacial ice
- Ocean surface temperature
- Thickness of the sea ice
- Surface elevation
- Surface roughness
- Distribution of NO_2.

In the present study, the prescribed ocean surface temperatures and the prescribed distribution and thickness of the sea ice are the observed annual means.

The albedo of the earth's surface is in part prescribed and in part diagnostically determined. Ice-free ocean has a prescribed albedo of 0.07 and the sea ice and snow-free glacial ice have prescribed albedos of 0.40. The albedo of the predicted snow cover is a diagnostic variable, taken as 0.65 when the snow is colder than its melting temperature and as 0.50 when the snow is at its melting temperature.

The albedo of the ice-free and snow-free land is also diagnostically determined. It is made to vary linearly from an albedo of 0.28 where the evapotranspiration coefficient is zero, to an albedo of 0.12 where the evapotranspiration coefficient is unity. The evapotranspiration coefficient, the ratio of the actual to the potential evapotranspiration, is in turn made to vary with the amount of water in the soil, from zero where there is no evapotranspiration-available soil water, to unity where the soil water exceeds one-third of the field capacity (which is taken as 15 gm of water cm^{-2}). It is through this dependence of both the evapotranspiration and the

surface albedo on the amount of water in the soil that we have parameterized the interaction of the atmospheric circulation and the surface vegetation.

d. Vertical structure

Fig. 1 shows the vertical structure of the model. The lower boundary follows the earth's surface elevation (the topography) which is prescribed. The upper boundary is the 1 mb pressure surface, which is the approximate pressure of the observed stratopause. The model atmosphere is divided into 12 layers and the boundaries of these layers follow the coordinate surfaces of a generalized σ-coordinate. From 1 to 100 mb there are seven layers, each of which has equal thickness in $\log p$ and is bounded by constant pressure surfaces. From 100 mb to the earth's surface the mass is divided into five layers in the ratios 1:2:2:2:2. The dashed lines in the figure show the levels at which the prognostic variables of horizontal velocity, temperature, ozone and water vapor are carried for each of the twelve layers. These levels are centered in $\log p$ for the layers above 100 mb, and are centered in p for the layers below 100 mb.

The uppermost layer of the model, called the "sponge layer", has a damping term designed to absorb upward propagating wave energy and thus prevent a spurious reflection of wave energy at the upper boundary.

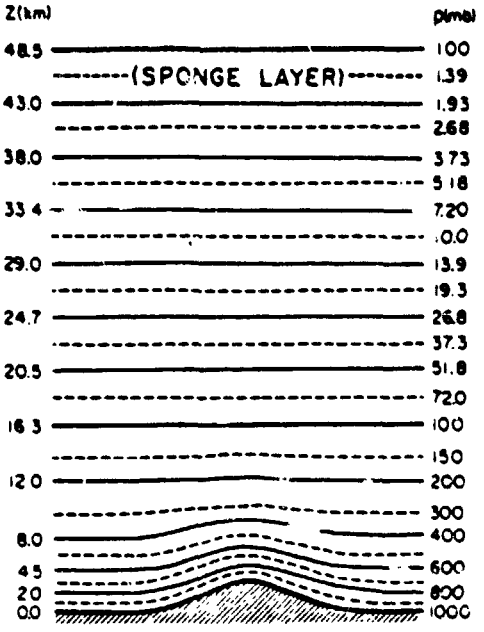


FIG. 1. Vertical structure of the model. The prognostic variables of horizontal velocity, temperature, water vapor and ozone are carried on the σ-surfaces shown by the dashed lines.

e. Horizontal coordinates, grid size and time step

The horizontal coordinates are longitude and latitude. The grid size is 5° of longitude and 4° of latitude, which makes the two grid distances equal at 37° latitude. To avoid having to use an extremely short time step because the meridians converge toward the poles, a longitudinal averaging is done of selected terms in the prognostic equations (Schlesinger, 1976b). The basic time step is 6 min, but the source and sink terms and the vertical transports of water vapor and ozone use a time step of 30 min.

f. Mathematical aspects

Arakawa (1966, 1970) developed the principle that the statistical properties of the finite-difference solution of the governing equations of atmospheric motion will approximate those of the continuous solution if the important integral constraints are maintained. The integral constraints which are maintained in the present UCLA general circulation model (Arakawa and Lamb, 1977) are 1) conservation of total mass; 2) conservation of total kinetic energy during inertial processes; 3) conservation of enstrophy (mean-square vorticity) during vorticity advection by the nondivergent part of the horizontal velocity (on a plane); 4) the integral constraint on the pressure gradient force; 5) conservation of total energy during adiabatic and nondissipative processes; and 6) conservation of total entropy and total potential enthalpy during adiabatic processes.

In the present model the variables are distributed over the horizontal grid points according to the scheme which best simulates inertia-gravity waves and the geostrophic adjustment process (Arakawa and Lamb, 1977). Consequently, solutions with this scheme are free from the highly ageostrophic computational noise which commonly exists in solutions obtained with other schemes.

The vertical difference scheme, when applied to levels equally spaced in $\log p$, guarantees that the vertical wavenumber of a planetary wave, for a given equivalent depth, is constant in height for an isothermal atmosphere, just as it is in the continuous case (Arakawa and Mintz, 1974). This is one of the important properties of the model, namely, that there is little or no spurious internal reflection of vertically propagating wave energy due to the discretization.

g. Initialization

For the present study, the prognostic variables of the model were initialized in the following way. For the σ-surfaces below 100 mb, the horizontal velocity, temperature and water vapor mixing ratio were initialized by vertical interpolation from a day at the beginning of December which was the

result of a long integration that had been made with an earlier UCLA three-level tropospheric general circulation model. The initial surface pressure, ground temperature, soil water storage and mass of snow on the ground were also taken from the three-level model results for that day. (The initial state for the integration of the three-level model was one in which the temperature field was three-dimensionally isothermal, the velocity was zero everywhere, the surface pressure was everywhere equal to the global-mean surface pressure, and there was no water vapor in the air, no water stored in the soil and no snow on the ground.)

For the σ-surfaces above 100 mb, the initial temperature for this study was taken as varying only with pressure, using the Mid-Latitude Spring/Fall temperature profile in the *U. S. Standard Atmosphere Supplements (1966)*. The initial horizontal velocity above 100 mb was set equal to the geostrophic velocity, except at the equator, where it was set equal to the mean of the geostrophic velocities at 4°N and 4°S. The initial water vapor mixing ratio above 100 mb was taken as zero. The O_3 mixing ratio was initialized from the observed zonal-mean O_3 number density for December, based on the *Ozone Data for the World* (1960–72), and on the initial air density of the model. The initial sun declination and sun distance were taken for the beginning of December. The prescribed NO_2 profile was taken equal to the profile of McElroy *et al.* (1974).

Several times during the first 60 simulated days of integration, small programming and coding errors were detected and corrected. At these times the sun's declination and distance were set back some days (re-initialized), but no re-initialization was made of the prognostic variables.

By Day 60, however, it was apparent that the original formulation for the vertical ozone transport, which was the arithmetic-harmonic mean scheme, was producing a large, computationally false, downward transport of O_3 (Schlesinger, 1976a). To correct this computational error, the logarithmic-conserving scheme was introduced on Day 60, and the sun's declination and distance were set back, for the last time, to 20 December. The O_3 number density was re-initialized to the observed zonal mean for December with a lower bound of 0.04 μg g^{-1} taken for the re-initialized O_3 mixing ratios for the σ-levels below 400 mb. To make the simulated global photochemical O_3 source closer to the observed nominal surface O_3 sink given by Tiefenau and Fabian (1972) (5×10^8 ton year^{-1} or 15.85 ton s^{-1}), the magnitude of the prescribed NO_2 profile was reduced by 43.5%. As noted by McElroy *et al.* (1974) and Cunnold *et al.* (1975), such a reduction brings the prescribed NO_2 distribution into closer agreement with the observations of Ackerman and Muller (1972) and Murcray *et al.* (1974) in the altitude range between 15 and 30 km.

The first results of this final integration, covering the period from 20 December to 5 January, were presented at the U. S. Department of Transportation Fourth Conference on the Climatic Impact Assessment Program (Mintz and Schlesinger, 1975). Following that, the integration was continued until 31 January. We are showing the results for the last 20 days of January, even though the O_3 distribution cannot reach its equilibrium state by the end of this period. For convenience, this period (12–31 January) will be referred to simply as January.

3. Simulated pressure, temperature, circulation and heating

a. Sea level pressure

The average January sea level pressure simulated by the model is shown in Fig. 2, together with the observed climatological January sea level pressure. We see that the model has fairly successfully simulated the general position, shape and strength of the subtropical oceanic high-pressure centers near 30° latitude in both hemispheres, and also the equatorial low-pressure zone and the low-pressure centers over the Southern Hemisphere continents. The location and configuration of the Icelandic and Aleutian lows are also simulated fairly well, inasmuch as the differences that are seen here between the simulation for a single January and the observed long-period normal for January are not larger than the differences that are observed from one January to another (see, e.g., Lamb 1972, pp. 426–427, and Bjerknes, 1975, p. 597). However, the central pressures in the simulated lows are lower than those observed, with the exception of some extreme years. The model has also produced a circumpolar low-pressure belt in the Southern Hemisphere, but its location near 55°S is about 10° of latitude too far from the pole and its pressure is about 10 mb too high. Along the Antarctic coast, the simulation shows two high-pressure centers, one over the Ross Sea and the other over the Weddell Sea, which are not found in the observed normal field. A more prominent shortcoming of the simulation is the excessively high pressure over the Arctic region, with a pressure of 1028 mb at the North Pole compared with the observed normal of 1016 mb. This excessive sea level pressure over the Arctic is accompanied by the failure to properly simulate the observed Siberian high-pressure center; so that the pressure gradient and the geostrophic wind have the wrong direction over northern Siberia, although the pressure gradients and the directions of the geostrophic wind are correct over eastern, southern and western Siberia.

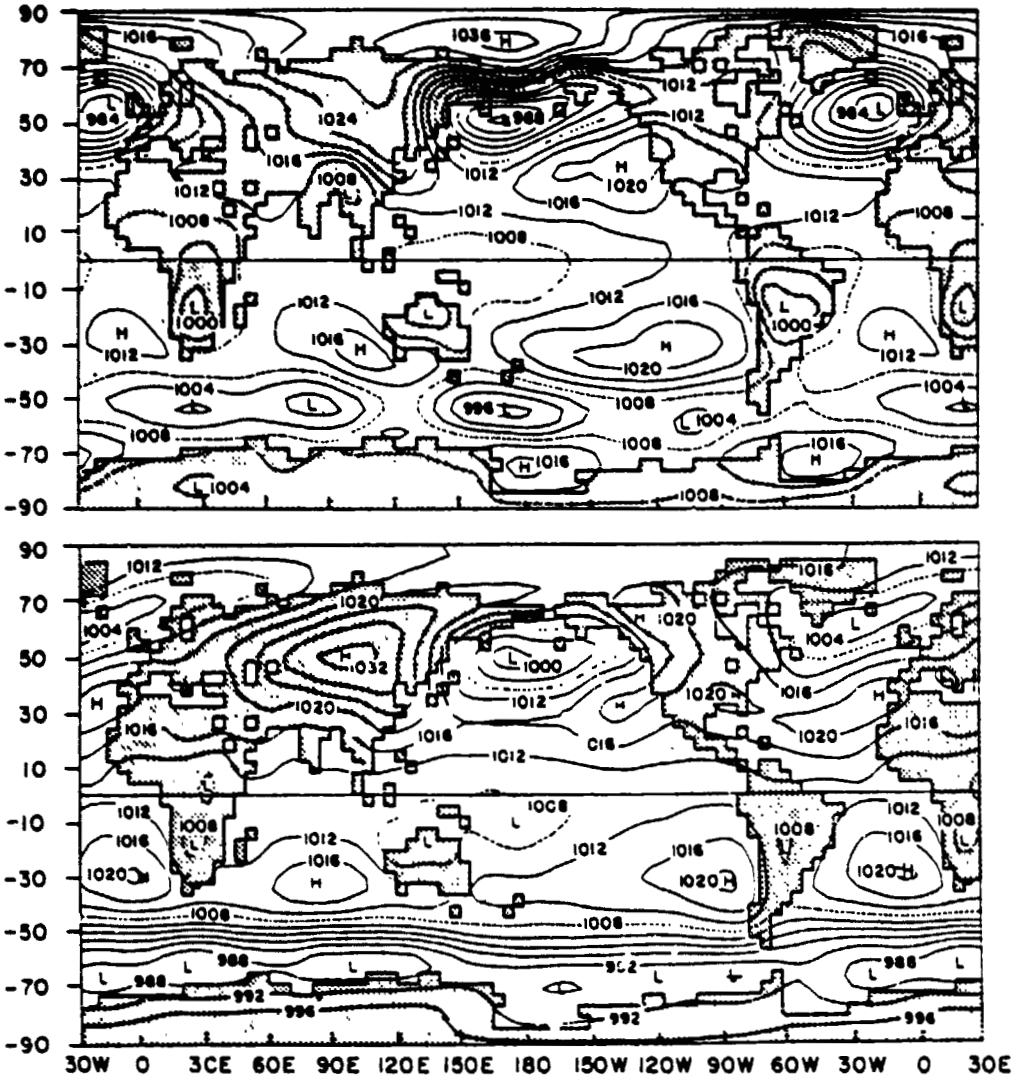


FIG. 2. Sea level pressure (mb). Top: Simulation, averaged for the last 20 days of the integration (12–31 January). Bottom: Observed normal for January [from Crutcher and Meserve (1970) and Taljaard *et al.*, (1969), as summarized by Schutz and Gates (1971)].

b. Geopotential height

The simulated January 500 mb geopotential height distribution is shown in Fig. 3, together with the observed January normal. In the Northern Hemisphere, the model has correctly simulated the positions and intensities of the trough near Japan, the ridge near the west coast of North America and the low center near Baffin Island. However, the trough observed over the east coast of North America and the ridge over the eastern North Atlantic are found about 40° too far east in the simulation, and the weak trough that is observed over eastern Europe is nonexistent in the simulation. Thus, while extratropical planetary wavenumbers 1, 2 and 3 are evident in the observed field, only wavenumbers 1 and 2 are seen in the simulation. At the North Pole the simulated geopotential height is about 150 m too high. In the Southern Hemisphere, the observed geopotential pattern is fairly well simulated, except that the circumpolar low-pressure trough near 60°S is about 10° of latitude too far from the pole and the geopotential height, in and south of the trough, is about 200 m too high.

The simulated and observed distributions of January 100 mb geopotential height are shown in Fig. 4. At this level, the Northern Hemisphere planetary wavenumber 3 is weakly developed in the observed

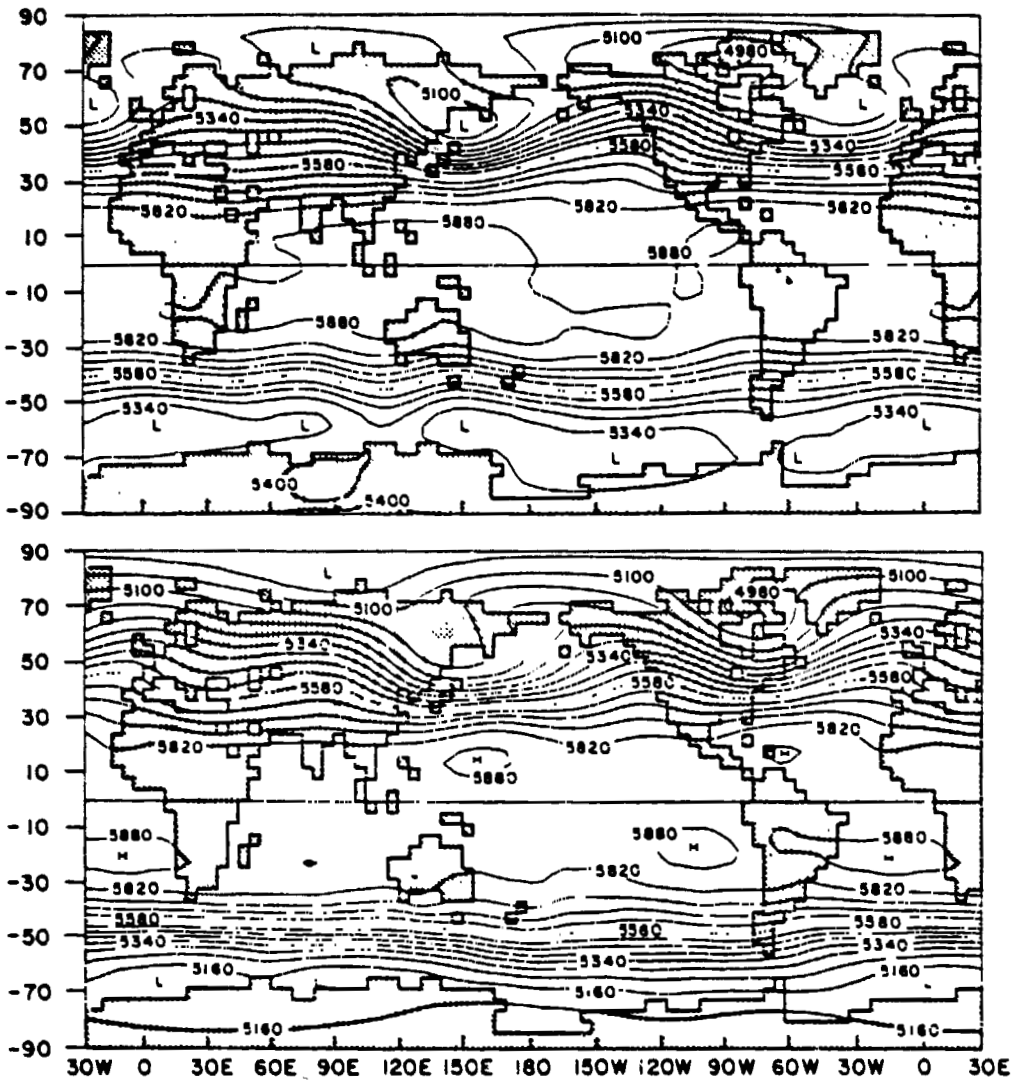


FIG. 3. 500 mb geopotential height (m). Top: Simulation, averaged for the last 20 days of the integration (12–31 January). Bottom: Observed normal for January [from Crutcher and Meserve (1970) and Taljaard *et al.*, (1969)].

field; and it does not appear at all in the simulation. The simulated 100 mb geopotential heights are about 600 m too low over the polar regions in both hemispheres. The simulated and the observed Southern Hemisphere circumpolar low-pressure troughs both show that the geostrophic winds are easterly over the South Pole at this level, as at 500 mb.

The simulated global distribution of January 50 mb geopotential height is shown in Fig. 5, together with the observed January distribution for the Northern Hemisphere. At this level both the simulated and the observed fields show a wavenumber 2 pattern in the Northern Hemisphere. The simulated geopotential height is too low (by about 900 m) in the North Polar region.

The simulated global distribution of January 10 mb geopotential height is shown in Fig. 6, together with the observed January distribution for the Northern Hemisphere. The model reproduces the high-pressure center that is observed over the North Pacific and shows the wavenumbers 1 and 2 of the observed field, but the simulated geopotential height in the North Polar region is about 2000 m too low.

c. Zonally averaged temperature

Fig. 7 shows a meridional cross section of the simulated zonally averaged January temperature, together with the observed normal temperature for December–February. In the source that we have

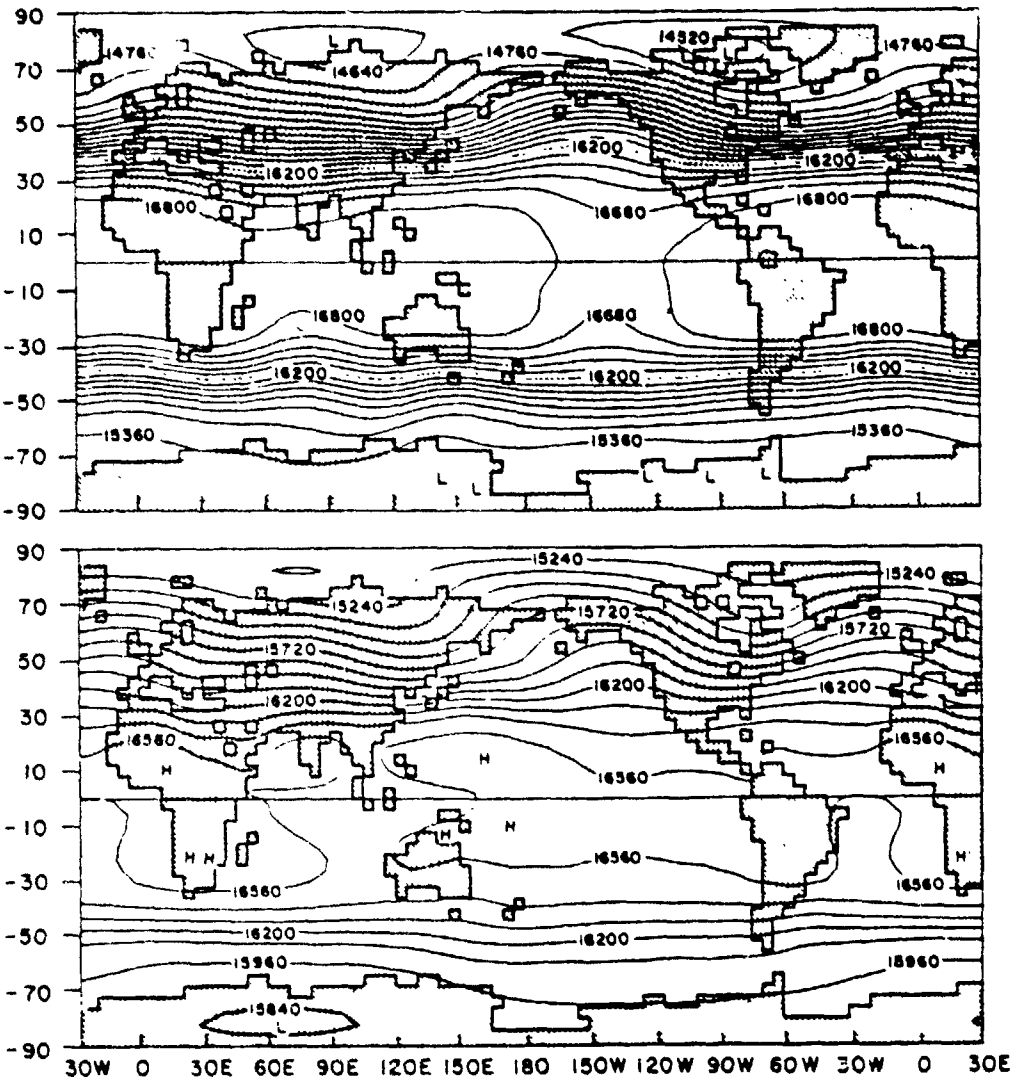


FIG. 4. 100 mb geopotential height (m). Top: Simulation, averaged for the last 20 days of the integration (12–31 January). Bottom: Observed normal for January [from Crutcher and Meserve (1970) and Taljaard *et al.*, (1969)].

used, the observed field is shown only up to the 10 mb (~31 km) level, and this level is indicated in the simulated field by a thin horizontal line.

Comparison of the simulated and observed fields shows that the zonal-mean tropospheric temperature field is fairly well reproduced by the model. At the equator, the tropopause temperature and the height of the tropopause (within the vertical resolution of the model) are well simulated. In the tropical latitudes, the temperatures of the lower and middle stratosphere are also well simulated. The Northern Hemisphere middle-latitude temperature maximum at the tropopause level, which is observed at about 55°N, is also simulated by the model, although it is about 10° of latitude too far from the pole and about 5°C too cold. In the Southern Hemisphere, the simulated temperature maximum at the tropopause level is about 20° of latitude too far from the pole and is about 15°C too cold. A more important shortcoming of the simulation, however, is that in the polar regions of both hemispheres the simulated temperatures in the upper troposphere and the lower and middle statosphere are 15–25°C colder than the observed temperatures. This shortcoming of the simulation was already seen in Figs. 4, 5 and 6, where we noted that the geopotential heights of the 100, 50 and 10 mb pressure surfaces were too low in the polar regions. The excessively cold air in the polar regions of the upper troposphere and the lower and middle stratosphere,

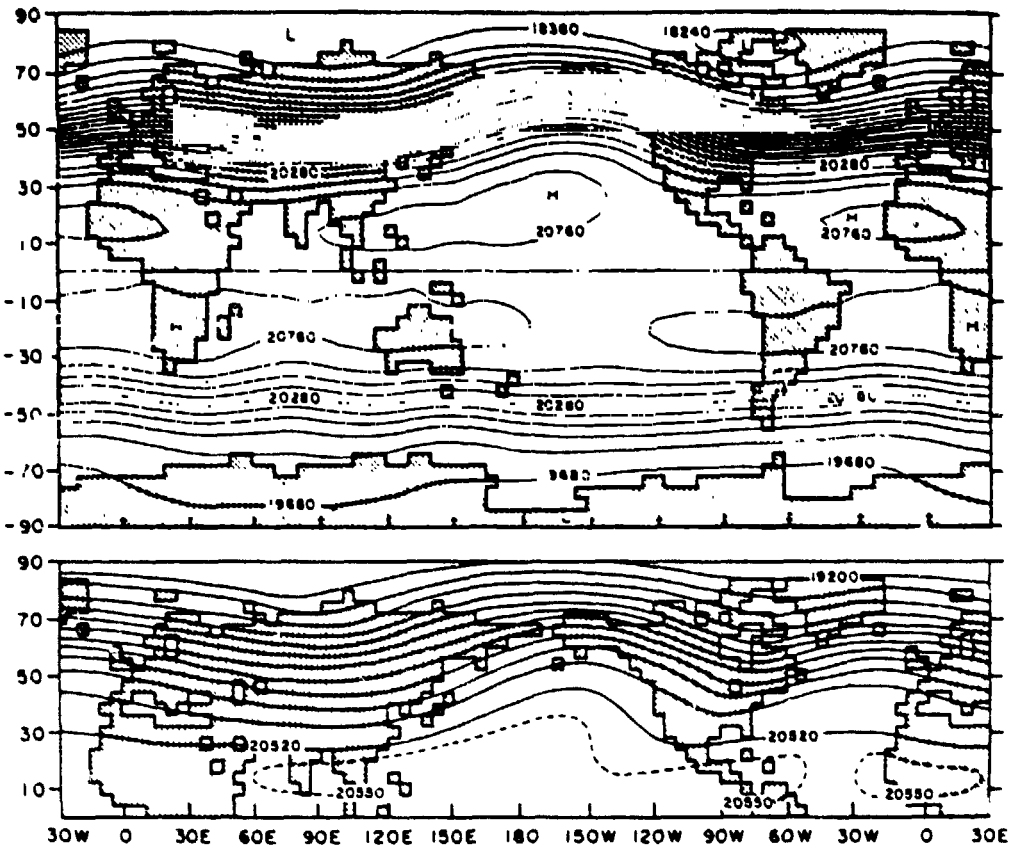


FIG. 5. 50 mb geopotential height (m) Top: Simulation, averaged for the last 20 days of the integration (12–31 January). Bottom: Observed January 12-year mean (from CIAP, 1974).

and the excessively high sea level pressure in the polar regions, seem to be common deficiencies of many atmospheric general circulation models whose upper boundary is higher than the tropopause level and in which the temperature field is not empirically constrained (Somerville *et al.*, 1974, Manabe and Mahlman, 1976; Washington *et al.*, 1977).

Although the simulated temperatures near the poles are too cold in the upper troposphere and lower and middle stratosphere, the pole-to-pole temperature distribution in the uppermost stratosphere is very close to what is observed, at this time of the year, in the period that precedes sudden warming, as shown by Barnett (1974).

d. Zonally averaged zonal wind

Fig. 8 shows the simulated zonally averaged zonal wind component for January, together with the observed normal for December–February. In the troposphere, the model correctly simulates the positions of the two subtropical west wind maxima, although their intensities are too large by about 20 m s^{-1}. The tropical easterlies are also fairly well simulated, including the fact that near the earth's surface the easterlies are stronger on the winter side of the equator, whereas in the middle and upper troposphere and the stratosphere, they are stronger on the summer side of the equator. In the polar regions of the troposphere and lower stratosphere, in both hemispheres, the simulation shows easterly zonal winds. In the south polar regions, the observed geostrophic winds, as shown in Figs. 3 and 4, are also easterly. In the north polar region, almost all analyses show that in the middle and upper troposphere and the stratosphere the observed normal January flow is westerly.

In the stratosphere, the model produces a polar-night westerly jet and a summer hemisphere tropical easterly jet. The simulated polar-night jet is located about 10° of latitude too far from the pole and is much too strong (at the 10 mb level it is about twice the speed of the observed jet). In addition, the stratospheric polar-night jet is not well separated from the tropospheric subtropical west wind maximum. The simulated stratospheric easterly winds are very similar to the observed ones, as shown by Murgatroyd (1969), but the simulated Southern Hemisphere west winds extend much too high into the stratosphere. All of these deficiencies, of course,

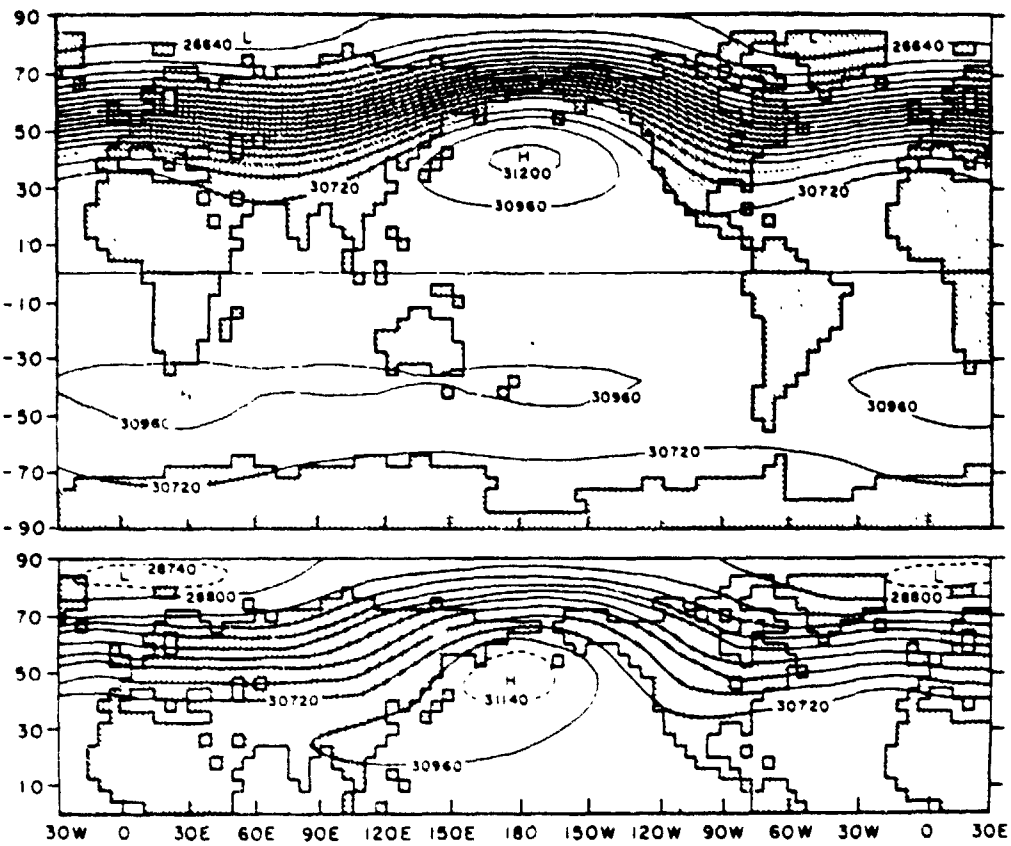


FIG. 6. 10 mb geopotential height (m). Top: Simulation, averaged for the last 20 days of the integration (12–31 January). Bottom: Observed January 9-year mean (from CIAP, 1974).

are geostrophically related to the excessively cold simulated temperatures in the upper troposphere and lower and middle stratosphere, in the polar regions of both hemispheres.

e. Mean-meridional mass circulation

The simulated January mean-meridional mass circulation is shown in Fig. 9, together with the observed December–February circulation. The simulated and the observed fields show the tropospheric tropical Hadley cells, with the Northern Hemisphere (winter) cell broader and more intense and extending across the equator, the tropospheric middle latitude Ferrel cells, and the secondary direct cells poleward of the Ferrel cells. The simulation has an additional indirect cell in the north polar region.

In the stratosphere, the dominant circulation feature in the winter hemisphere is the upward current whose axis crosses the 100 mb level near 60°N and the 10 mb level near 70°N, then turns equatorward and, merging with a current from the south, descends across the 10 and 100 mb levels at about 40°N, where it joins the converging upper branches of the tropospheric Hadley and Ferrel cells and continues to the ground. This circulation feature resembles the one observed by Newell *et al.* (1973) for a winter season. The axis of the descending current, at about 40°N latitude at the tropopause level, is close to the temperature maximum at the tropopause level shown in Fig. 7.

The boundary between the two Hadley cells is the mean intertropical convergence zone (ITCZ), which is about 5° of latitude too far south in the simulation. Part of the upward current in the ITCZ, reinforced by part of the upward branch of the Southern Hemisphere Ferrel circulation, penetrates through the lower and middle stratosphere as a broad current and then crosses the equator in the upper stratosphere to join the current from the north; the combined stream, as already indicated, then descends and returns to the troposphere near 40°N.

f. Zonally averaged heating

Figs. 10, 11 and 12 show the simulated zonally averaged heating for January. The heating rates are shown in terms of an equivalent time rate of change

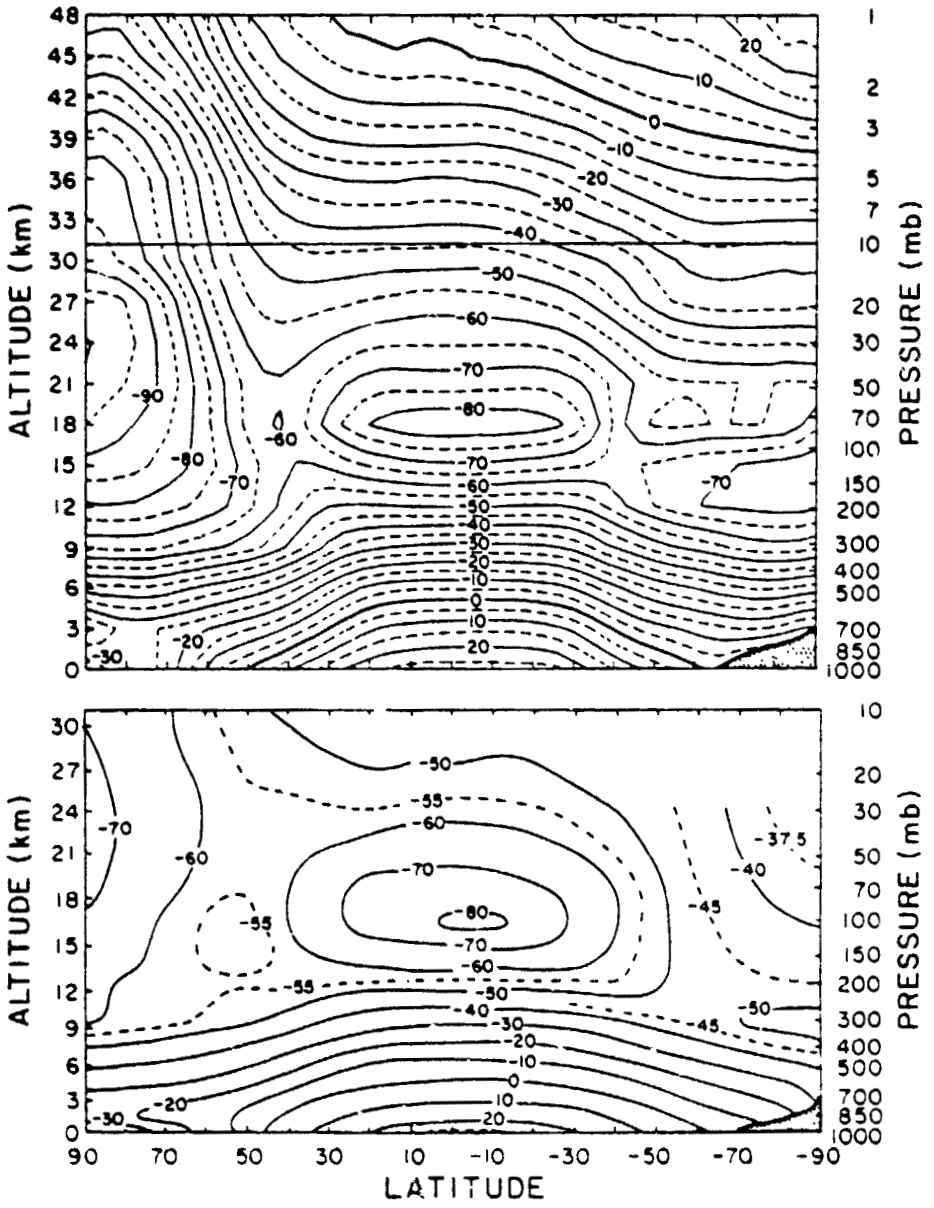


FIG. 7. Zonally averaged temperature (°C). Top Simulation, averaged for the last 20 days of the integration (12–31 January). Bottom: Observed normal for December–February (from Newell *et al*., 1972).

of the temperature of the air. The solid line interval is 0.5°C day^{-1} and the regions of cooling (negative heating) are shaded.

The top panel of Fig. 10 shows the heating by cumulus convection, which is the parameterized subgrid-scale release of latent heat and its upward convective transfer. Most of this heating is in the tropics, where it reaches to the tropopause level.

The bottom panel of Fig. 10 shows the large-scale (i.e., grid-scale) condensation heating. Because this requires large-scale saturation, almost all of it is in extratropical latitudes, where there is very little cumulus convection. Large-scale saturation cannot occur in the presence of cumulus convection, except at the uppermost detrainment level, because the large-scale relative humidity is kept below saturation by the subgrid-scale sinking motions that surround the cumulus cloud towers. The large-scale condensation, moreover, is mainly confined to the lower troposphere because of the strong dependence of the saturation vapor pressure on temperature.

However, there is a small amount of large-scale condensation in the tropics (producing zonal-mean

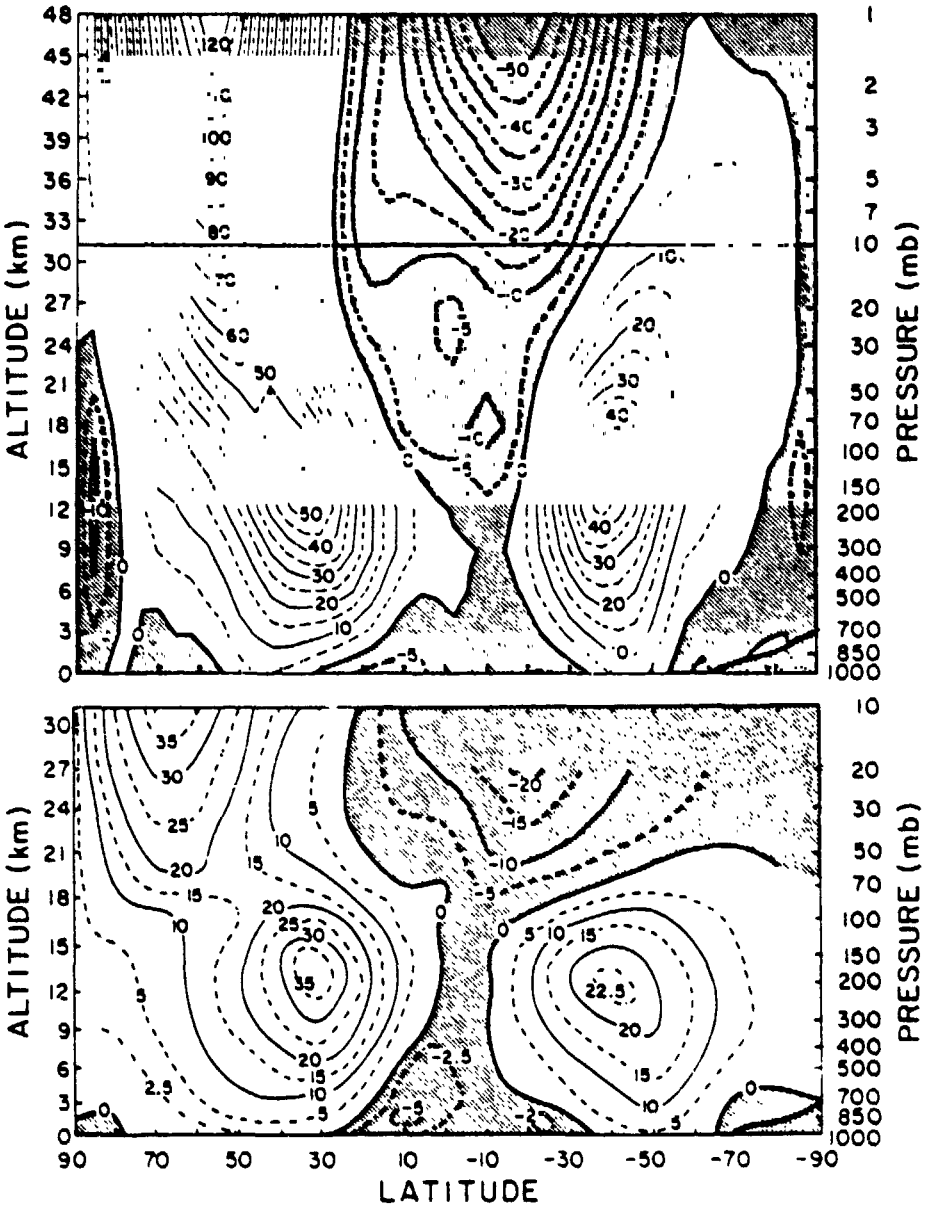


FIG. 8. Zonally averaged zonal velocity (m s^{-1}). Top: Simulation, averaged for the last 20 days of the integration (12–31 January). Bottom: Observed normal for December–February (from Newell *et al.*, 1972). Easterlies are shaded.

heating of less than 0.25°C day^{-1}). A very small amount of this is in the upper troposphere, where large-scale saturation and condensation are produced by the detrainment of water vapor from the tops of the tall cumulus clouds and by the upward large-scale motion which reaches, and even penetrates, the tropopause. It is the precipitation of this large-scale condensation, falling and evaporating into the air below, where the large-scale relative humidity is kept low by the cumulus convection, that produces the region of small net evaporational cooling, shown by the shaded region in the bottom panel of Fig. 10.

The top panel of Fig. 11 shows the solar radiational heating. In the model this heating depends on O_3 absorption from the upper boundary down to the 200 mb level and on water vapor and water cloud and ice cloud absorption from the 100 mb level down to the earth's surface. The O_3 absorption produces an increase of the heating with height. The absorption by the water substances produces a general decrease of the heating with height.

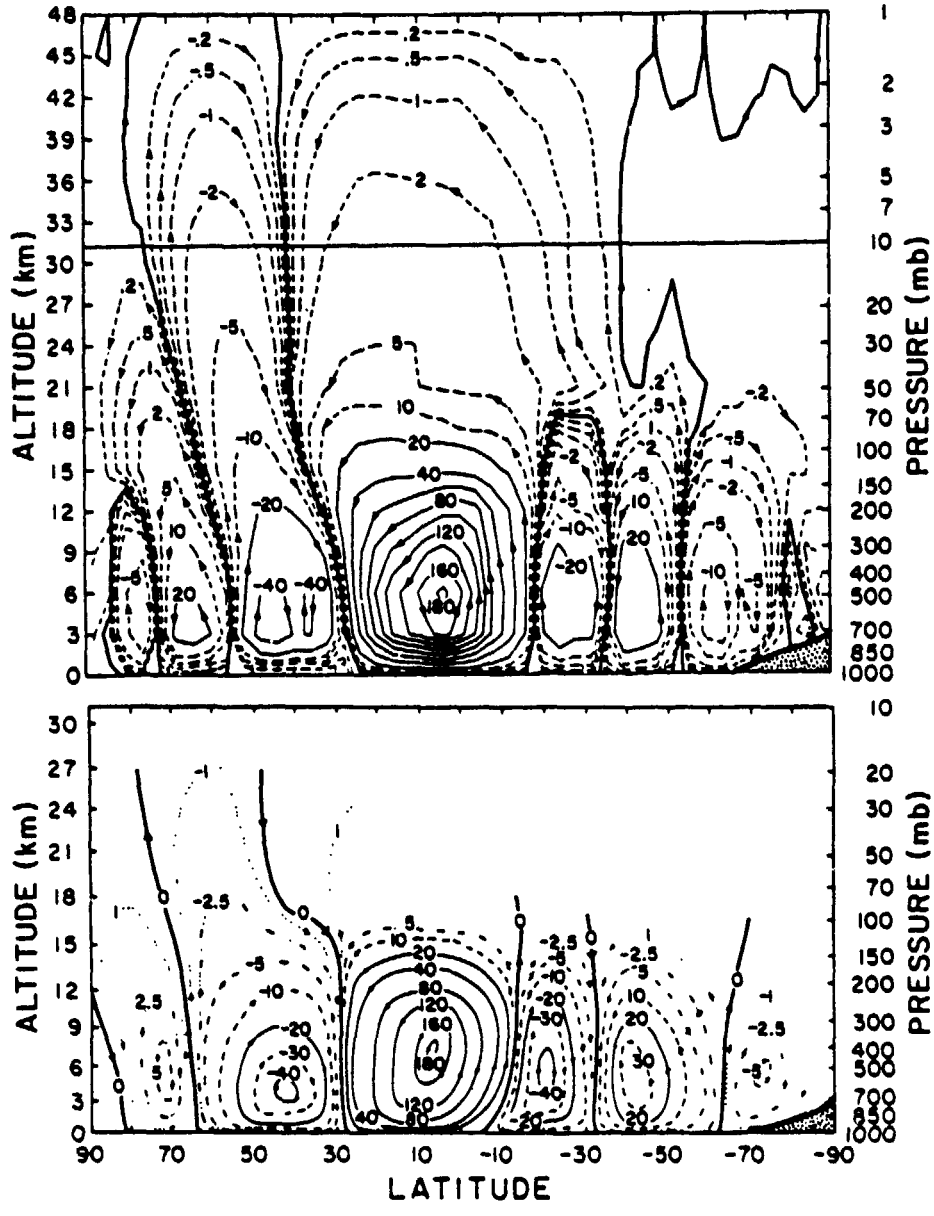


FIG. 9. Mean-meridional mass circulation (10^9 kg s^{-1}). Top: Simulation, averaged for the last 20 days of the integration (12–31 January). Bottom: Observed normal for December–February (from Newell *et al.*, 1972).

The bottom panel of Fig. 11 shows the longwave radiational heating. In the model, below 12 mb (30 km), this heating depends on ozone, carbon dioxide and water vapor and, below 100 mb, on water clouds and ice clouds as well. Above 12 mb, the longwave radiational heating follows the cooling rate approximation of Dickinson (1973). The simulation shows a longwave radiational cooling which decreases with height in the troposphere and increases with height in the stratosphere. It is only at and immediately above the tropical tropopause that the carbon dioxide and ozone absorption bands, in that very cold air, have a large enough direct energy exchange with the warm tropical ocean, and with the very warm midday tropical land surfaces, to result in a net longwave radiational heating of the air, as shown in the figure by the region which has no shading.

The top panel of Fig. 12 shows the total zonal-mean radiational heating, which is the sum of the two fields shown in Fig. 11. Over most of the domain, the total radiational heating is the small difference of two large terms of opposite sign. Within the troposphere, the lines of constant zonal-

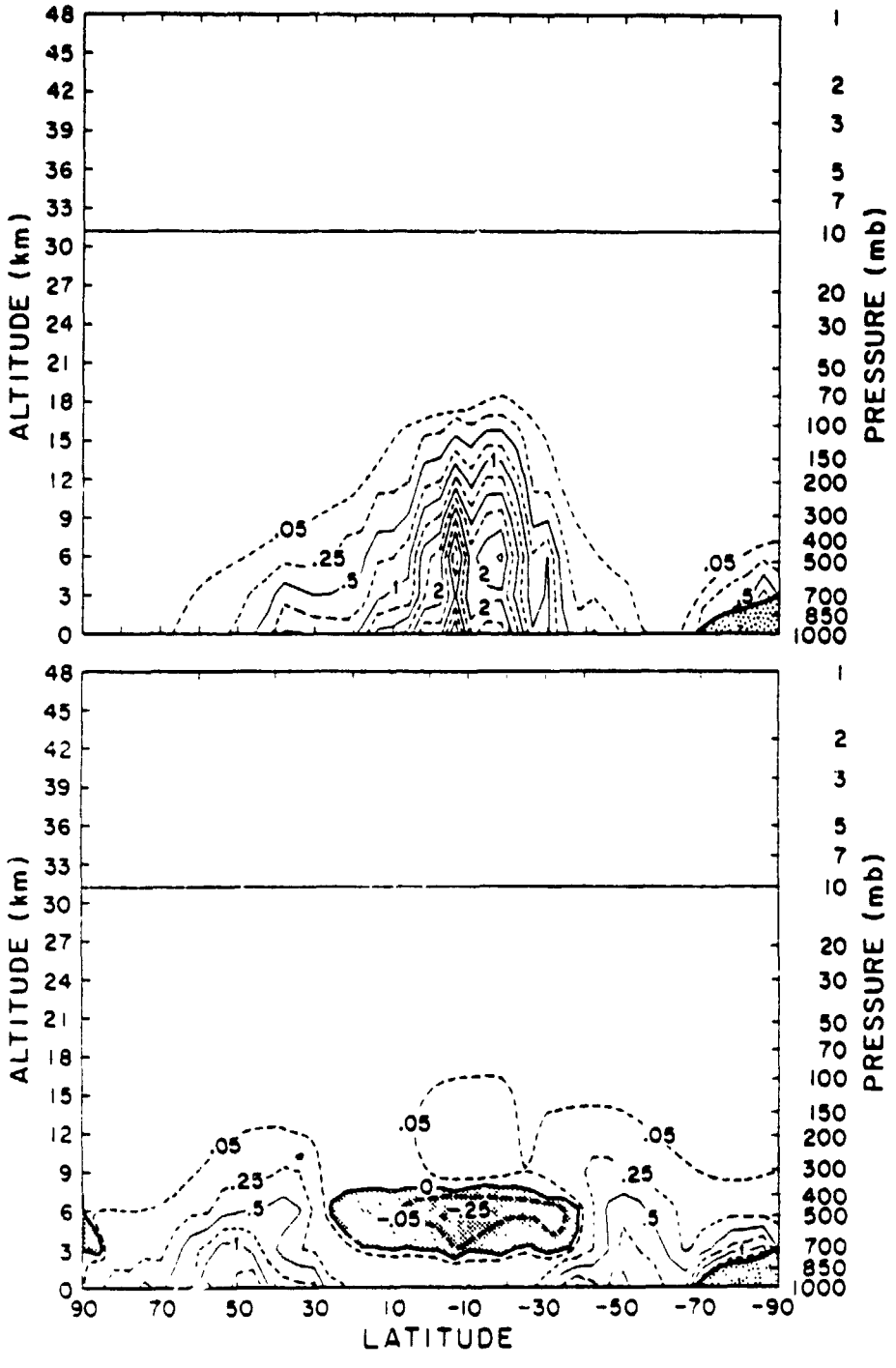


FIG. 10. Zonally averaged heating rate (°C day^{-1}) averaged for the last 20 days of the integration (12–31 January). Top: Heating by cumulus convection. Bottom: Heating by large-scale condensation. Negative heating (i.e., cooling) is shaded. Solid line contour interval is 0.5°C day^{-1}.

mean radiational heating are quasi-horizontal, so that the radiation, by itself, has very little influence on the zonal available potential energy of the troposphere. In the stratosphere, by contrast, there are large horizontal gradients of the radiational heating; how these affect the stratospheric zonal available potential energy we shall see presently.

The bottom panel of Fig. 12 shows the total zonal-mean heating of the atmosphere due to all diabatic processes: the solar and the longwave radiational heat-

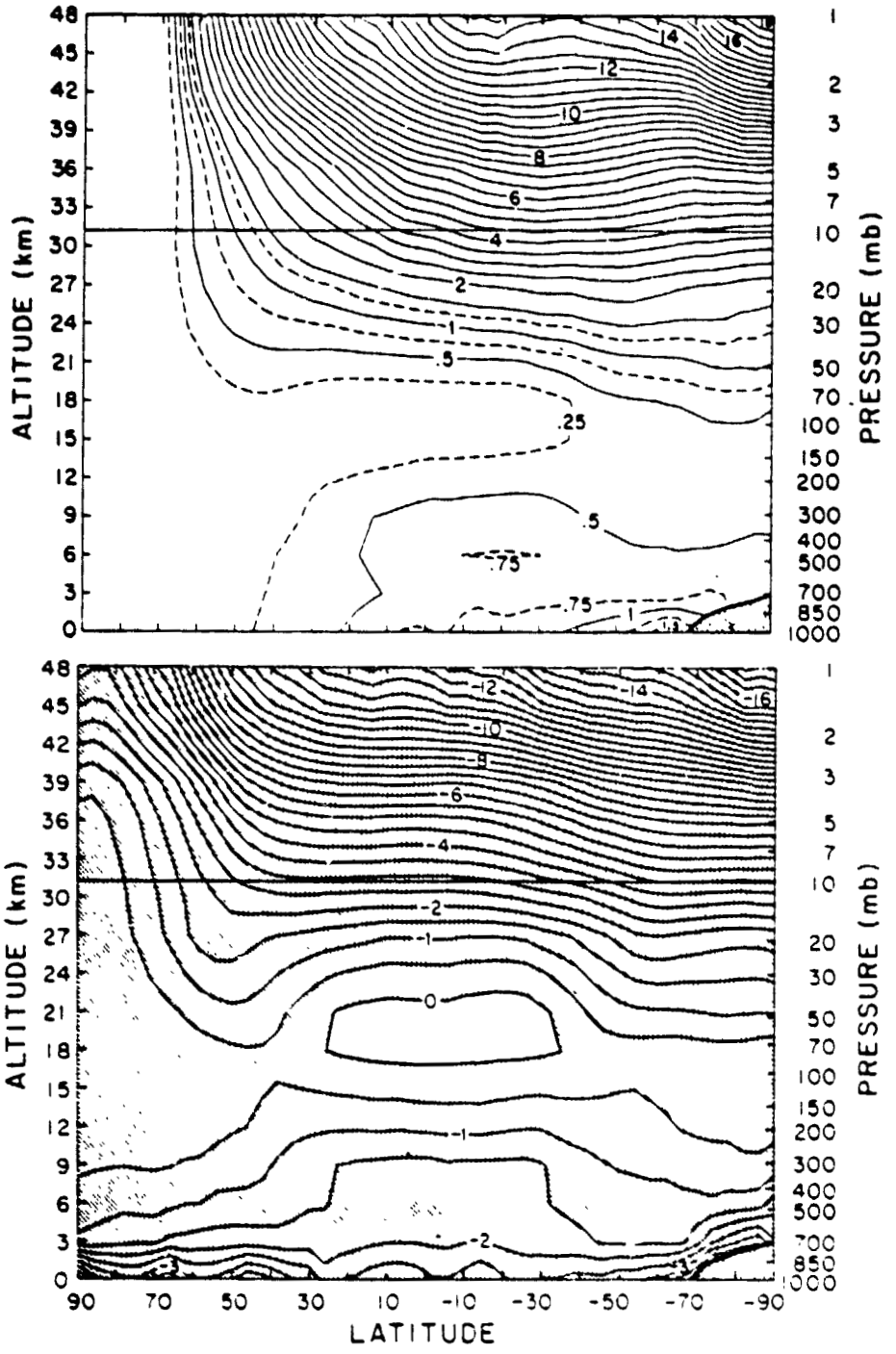


FIG. 11. Zonally averaged radiational heating rate (°C day^{-1}) averaged for the last 20 days of the integration (12–31 January). Top: Solar radiational heating. Bottom: Longwave radiational heating. Negative heating (i.e., cooling) is shaded. Solid line contour interval is 0.5°C day^{-1}.

ing shown in Fig. 11, the cumulus convection and the large-scale condensation heating shown in Fig. 10, and the parameterized boundary-layer heating (not shown in a separate panel) which produces the net heating of the lowest model layer in the subtropical and middle latitudes of the Northern Hemisphere. Also shown in the total heating field, in the uppermost stratosphere near the North Pole, is a subgrid-scale, dry-convective heat transfer from the second highest to the highest model layer.

The numerical simulation has provided us with the data for making a complete analysis of the

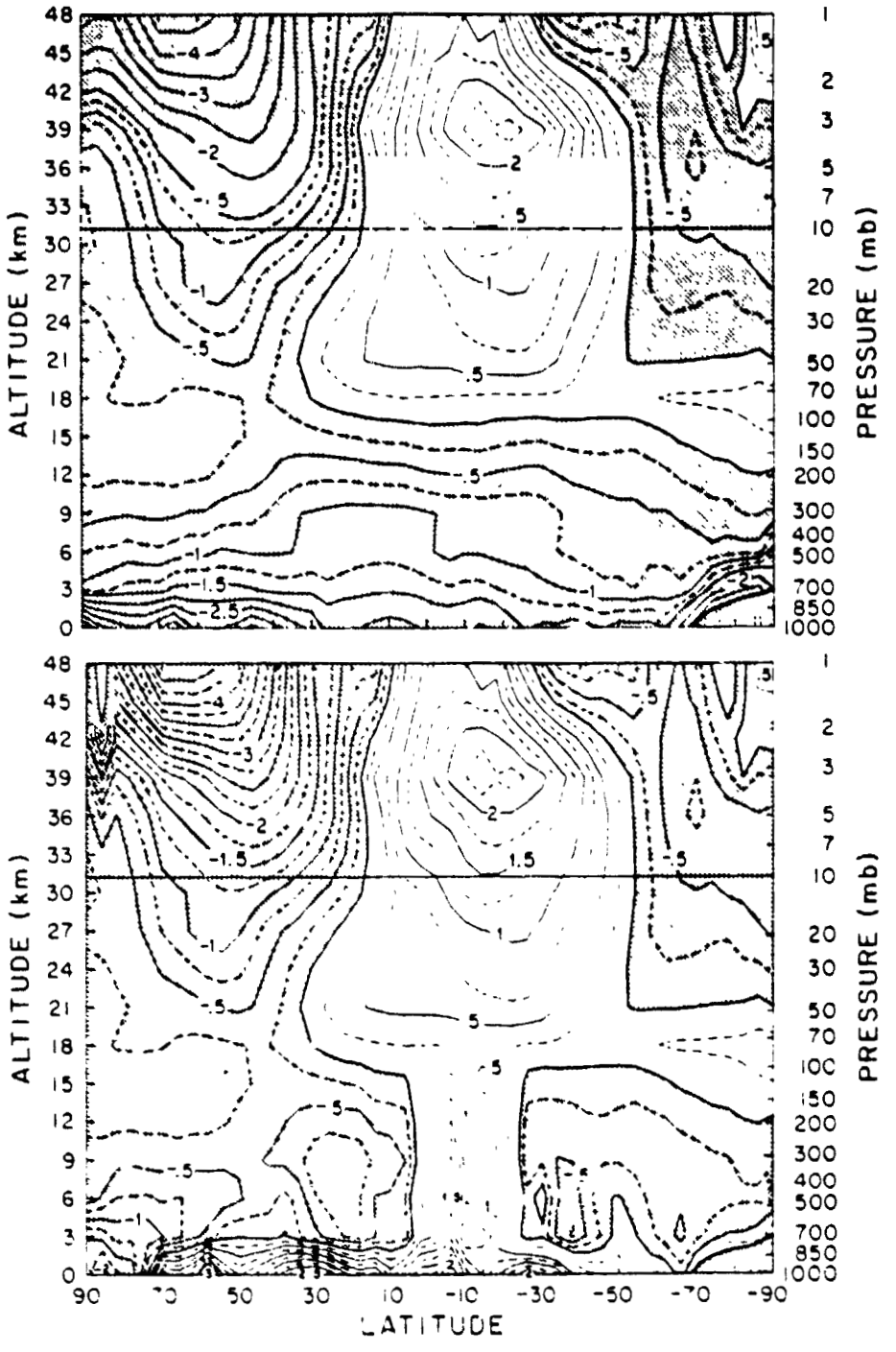


FIG. 12. Zonally averaged heating rate (°C day^{-1}) averaged for the last 20 days of the integration (12–31 January). Top: Total radiational heating. Bottom: Total heating of the atmosphere due to all diabatic processes. Negative heating (i.e., cooling) is shaded. Solid line contour interval is 0.5°C day^{-1}.

energetics of this atmospheric general circulation. That energetics involves many interactions between the zonal-mean fields and the zonal eddies, and its analysis is beyond the scope of this paper. Therefore, here we will only call attention to some relations among the zonal-mean fields themselves.

First, if we compare the zonal-mean heating and the zonal-mean temperature (as by superposing the bottom panel of Fig. 12 on the top panel of Fig. 7), we see that in the troposphere, except within the lowest model layer, there is a generation of zonal available potential energy by the heating of the air

near the equator, where the zonal-mean temperature is high, and cooling in all of the other latitudes, where the temperatures are lower. In the stratosphere, and especially in the lower stratosphere, except for the winter polar region, the opposite is true: there is a negative correlation between the heating and the temperature, so that zonal available potential energy is destroyed in the stratosphere.

Next, if we compare the zonal-mean temperature and the mean-meridional mass circulation (by superposing the top panel of Fig. 7 on the top panel of Fig. 9), we see that in the tropical troposphere there is a net conversion of zonal-mean available potential energy into zonal-mean kinetic energy due to the two Hadley cell circulations, in which the air that rises is warmer than the air that descends. In the stratosphere, by contrast, the ascending branches of the mean-meridional mass circulation are, on the whole, colder than the descending branches, so that zonal-mean kinetic energy is converted into zonal-mean available potential energy.

Finally, we compare the mean-meridional mass circulation with the zonal-mean heating field (by superposing the top panel of Fig. 9 on the bottom panel of Fig. 12). In the troposphere, the mean-meridional mass circulation has a positive feedback on the zonal-mean heating field in that the lower branches of the two Hadley cell circulations produce a convergence of the water vapor transport within the planetary boundary layer, which maintains the cumulus-convective heating (a convective instability of the second kind). In the stratosphere, except for the winter polar region, there is a negative feedback of the mean-meridional mass circulation on the heating field because the mean-meridional transport of ozone, in part, is responsible for the ozone distribution which contributes to the net radiational destruction of the zonal available potential energy.

4. Simulated ozone distribution

a. Total ozone

1) Synoptic behavior

Fig. 13 shows the simulated global distributions of total ozone (Ω), 500 mb geopotential height, and sea level pressure at 0000 GMT 22 January. A prominent characteristic of the synoptic Ω distribution is its large zonal variation, especially in the middle latitudes. The differences between adjacent maxima and minima along a latitude circle are of the order of several tens of Dobson units.

The 500 mb geopotential height and the sea level pressure fields show the transient wave cyclones. We see that the middle-latitude Ω maxima and minima coincide, respectively, with the trough and ridge lines of the 500 mb waves. For example, the Ω maximum west of France is located in the trough of a 500 mb wave, while the Ω minimum north of the Caspian Sea is located on the ridge of a 500 mb wave.

Fig. 14 shows the distribution of the total ozone at 30 h intervals to illustrate the general eastward movement of the Ω maxima and minima. The top panel is the same as in the previous figure (0000 GMT 22 January) and the middle and bottom panels are for 0600 GMT 23 January, and 1200 GMT 24 January. In the Northern Hemisphere, from 22 to 24 January the strong Ω maximum near Kamchatka moves eastward to almost the Aleutian Islands, the maximum near the Gulf of Alaska moves to the west coast of North America (and weakens), and the weak maximum over Alaska moves toward Victoria Island. Over Manchuria a new Ω maximum forms on 23 January (middle panel) and intensifies during the following day. From 22 to 24 January, the Ω maximum over Boston moves eastward to the central Atlantic, the maximum west of France moves southeastward to the Aegean, and the minimum north of the Caspian Sea moves to the east of the Aral Sea. Similar eastward movements of the Ω maxima and minima occur in the middle latitudes of the Southern Hemisphere.

Fig. 15 shows the simulated synoptic distributions of total ozone, 500 mb geopotential height and sea level pressure at the last time shown on the previous figure (1200 GMT 24 January). Again we see that the middle latitude Ω maxima and minima coincide, respectively, with the trough and ridge lines of the 500 mb waves. For example, the new Ω maximum which developed over Manchuria on the preceding day coincides with the newly developed 500 mb trough northwest of Japan, and is to the west of the new surface low that developed east of Japan. It is gratifying that both the amplitude of the zonal variation of total ozone and the phase relationship between the Ω maxima and minima and the troughs and ridges of the extratropical tropospheric waves are precisely what have long been observed in the real atmosphere, as was shown most clearly, for example, by Normand (1953). This correlation was also found in the O_3 tracer simulation study by Hunt and Manabe (1968).

2) January mean

The time-averaged distribution of total ozone for 12–31 January of the simulation is shown in Fig. 16, together with the observed January 10-year mean as derived from ground-based observations. In the simulation there are two Ω maxima in the Northern Hemisphere middle latitudes, a very intense one located just north of Japan and a weaker one over the central North Atlantic. In the Southern Hemisphere middle latitudes there are also two Ω

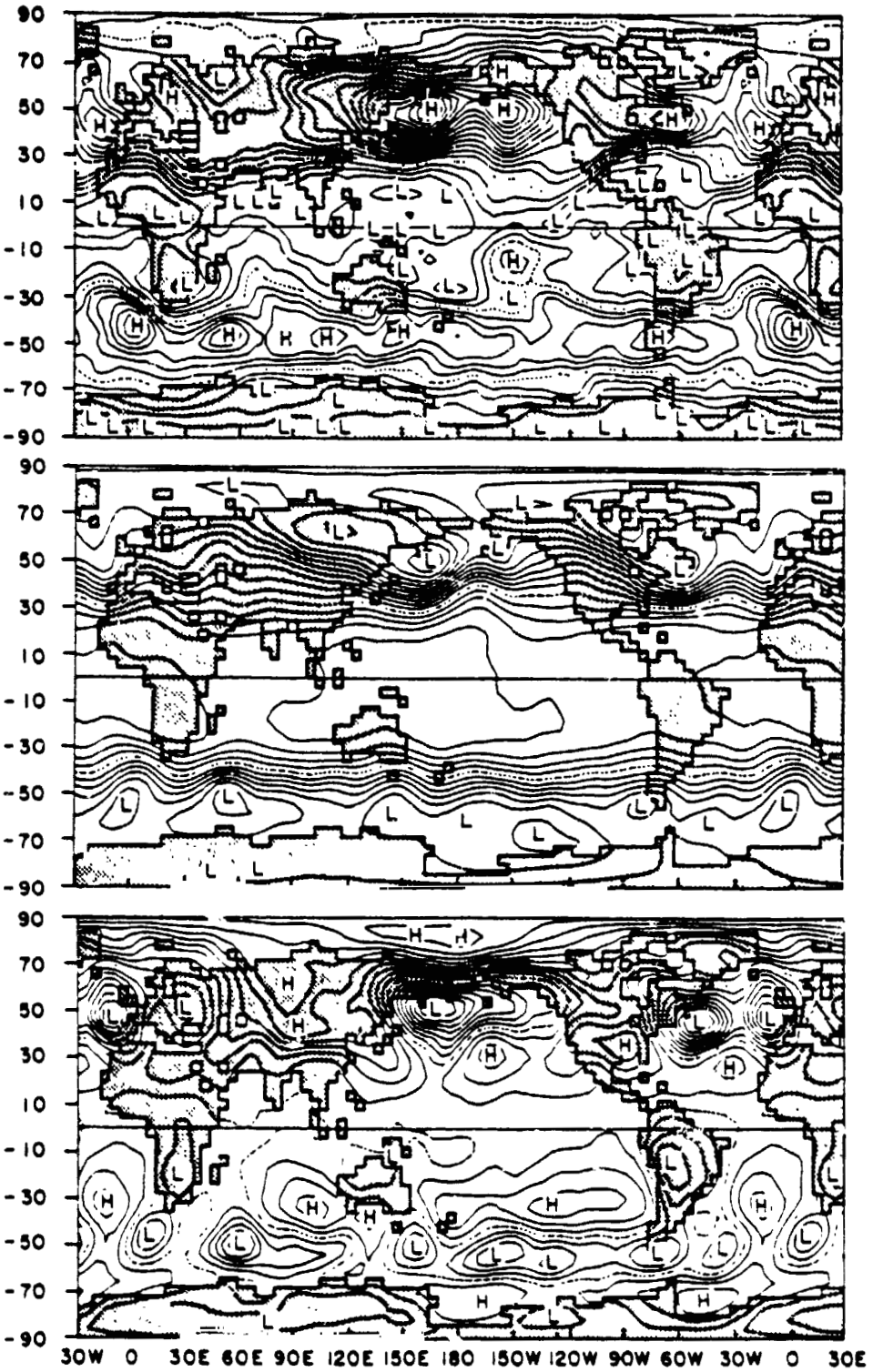


FIG. 13 Simulated synoptic distributions of total ozone, 500 mb geopotential height and sea level pressure, at 0000 GMT 22 January of the integration. Top: Total ozone, where the dashed contour line is 280 Dobson units and the contour interval is 10 D U Middle: 500 mb height, where the dashed contour line is 5580 m and the contour interval is 60 m. Bottom. Sea level pressure, where the dashed contour line is 1008 mb and the contour interval is 4 mb.

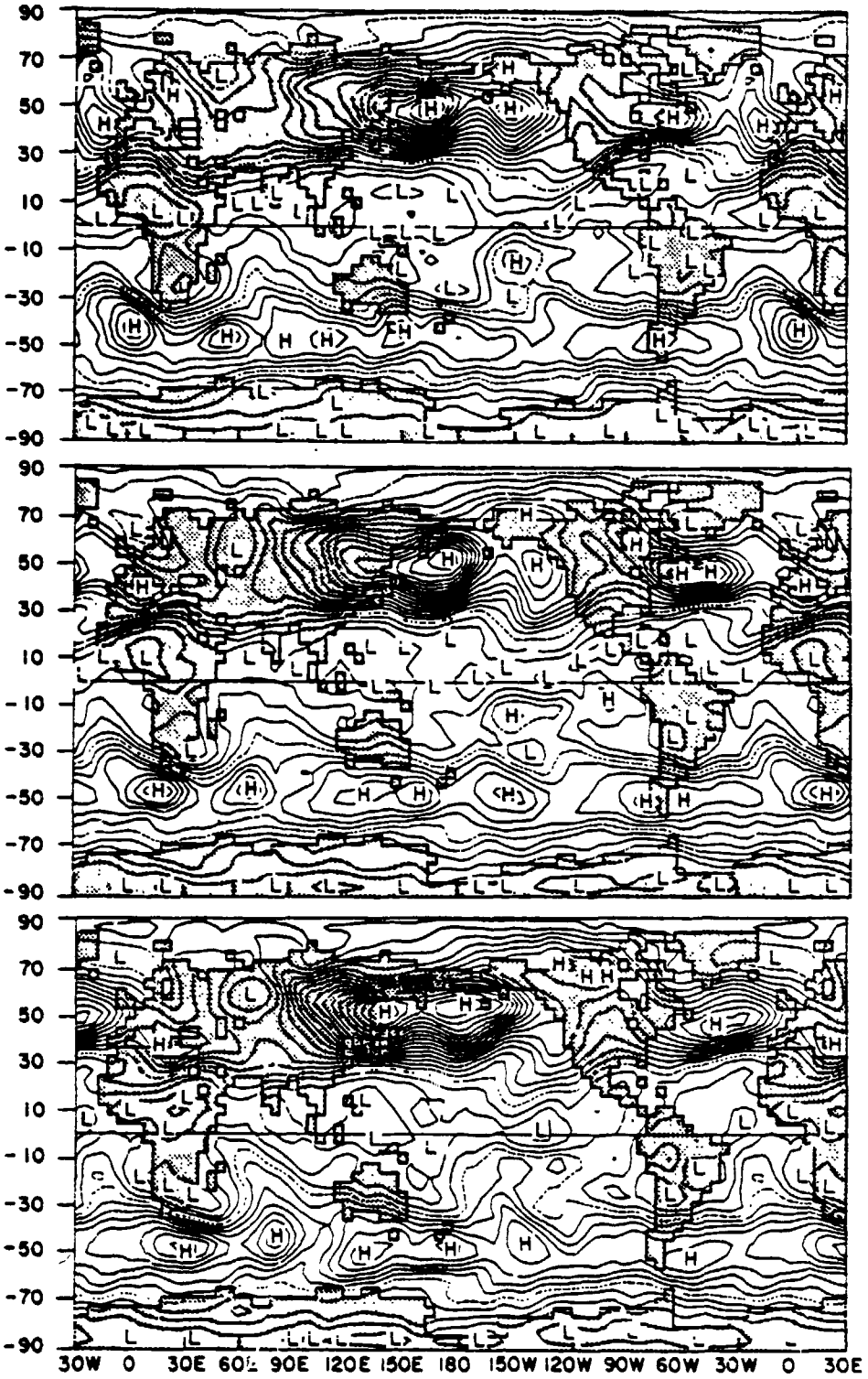


FIG. 14. Simulated synoptic distributions of total ozone at intervals of 30 h. Top: 0000 GMT 22 January Middle: 0600 GMT 23 January. Bottom: 1200 GMT 24 January. The dashed contour line is 280 Dobson units and the contour interval is 10 D.U.

maxima, one located south of Africa and another south of Tasmania. In the Southern Hemisphere tropics the simulated total ozone shows a wave-number 3 pattern, with the maxima positioned over the Indian, South Pacific and South Atlantic oceans.

The observed January-mean total ozone shows

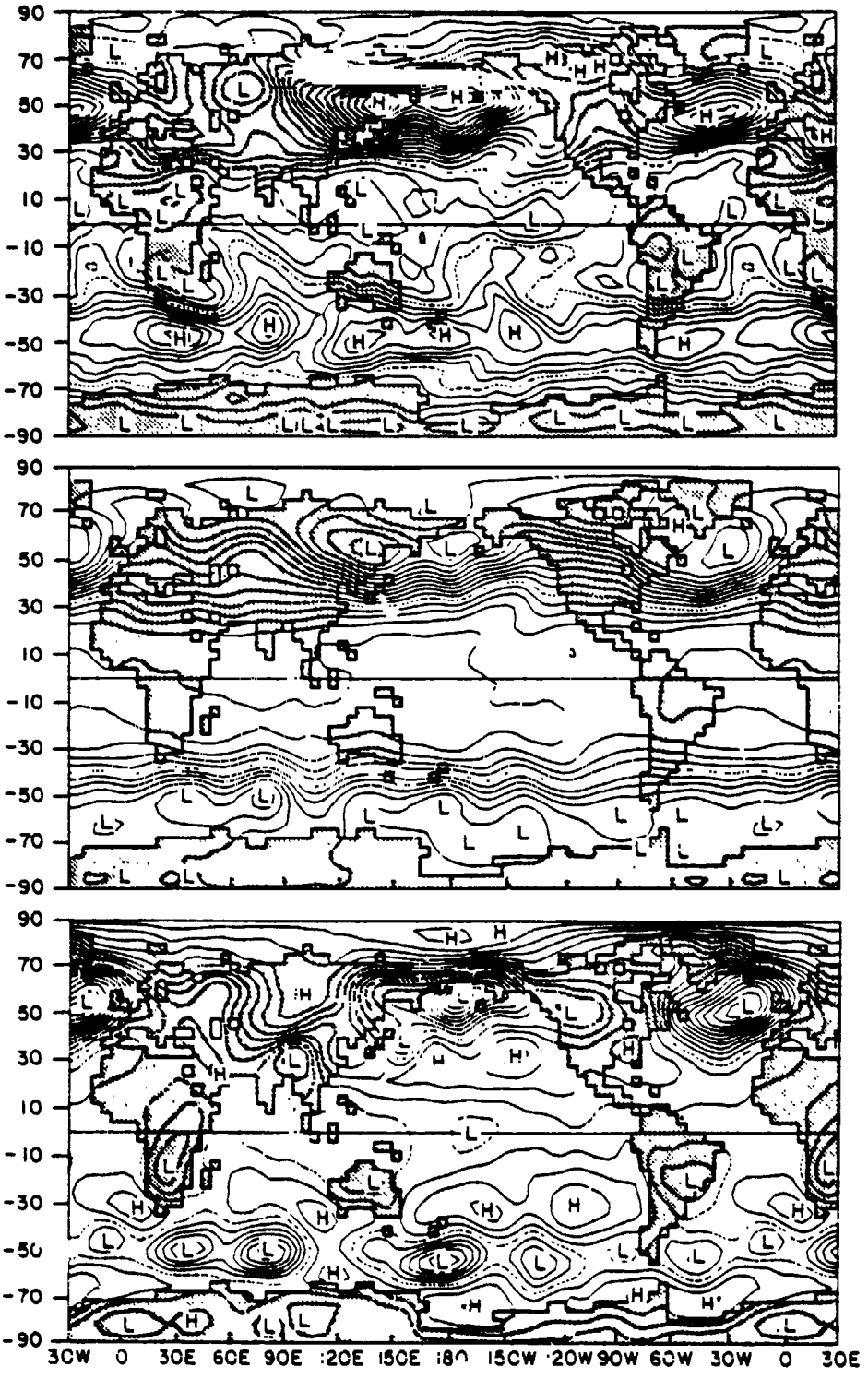


FIG. 15. As in Fig. 13, except 60 h later at 1200 GMT 24 January.

three Ω maxima in the Northern Hemisphere, one over Manchuria, one over Hudson's Bay and one over north-central Europe. In the Southern Hemisphere, only a single high-latitude maximum is seen at about 100°E and 55°S; and in the tropics the total ozone shows minima (and not maxima) over the three oceans.

If we compare the upper panel of Fig. 16 with the

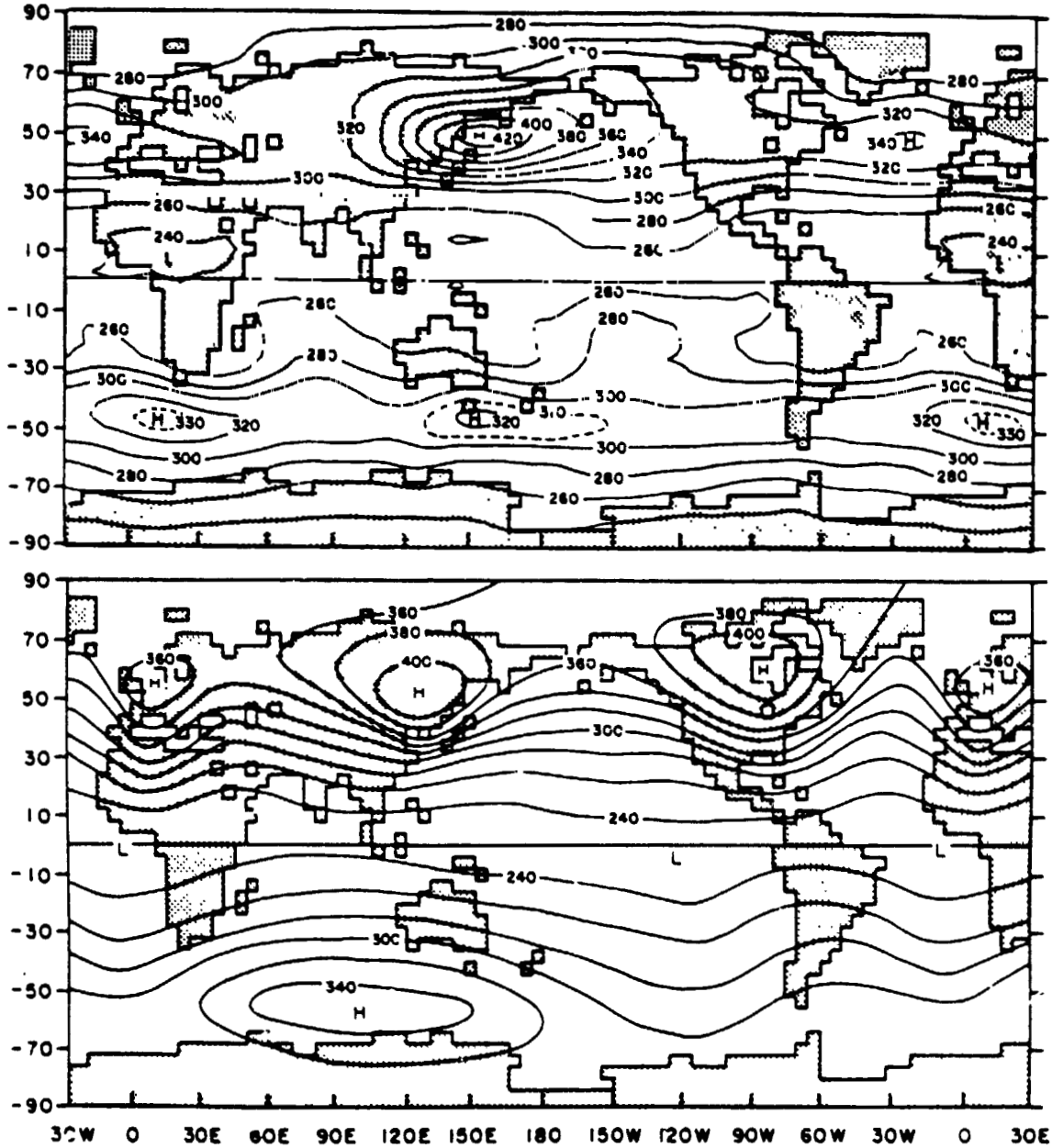


FIG. 16. Total ozone (Dobson units). Top: Simulation, averaged for the last 20 days of the integration (12–31 January). Bottom: Observed January 10-year mean (from London *et al.*, 1976).

upper panel of Fig. 3, and the lower panel of Fig. 16 with the lower panel of Fig. 3, we see that the simulated January total Ω distribution has about the same relationship to the simulated 500 mb height field as the observed Ω distribution has to the observed height field. That is, in both the simulated and observed fields the total Ω maxima are longitudinally positioned at the trough lines of the waves in the 500 mb height field. Moreover, just as the centers of the cyclonic circulation (the centers of maximum relative vorticity) are about 10° of latitude too far from the poles in the simulation, so also are the simulated Ω maxima about 10° of latitude too far from the poles.

Fig. 17 shows the zonally averaged initial total ozone distribution, the simulated zonally averaged January-mean distribution and the observed January-mean distribution. In the course of the simulation, the total ozone decreased in the high latitudes, especially in the Southern Hemisphere, and increased somewhat in the middle latitudes and the tropics, and the Ω maxima moved farther away from the poles.

We have already noted that both the longitudinal

and latitudinal positions of the individual Ω maxima have the same relationship to the 500 mb height field as those observed. Here we see that the zonally averaged total ozone is also properly related to the zonally averaged 500 mb geopotential height. In both hemispheres, the simulated zonally averaged 500 mb heights have their subpolar minima located about 10° of latitude too far from the poles, and the maxima in the zonally averaged total ozone are also located about 10° of latitude too far from the poles. It seems, therefore, that the dynamical and photochemical processes which relate the total ozone to the tropospheric circulation are operating correctly in the model, and that if the general circulation model can be made to produce a more nearly correct circulation field it would also produce a more nearly correct Ω field.

3) Correlations with tropospheric pressure

We have seen, in both the synoptic and time-averaged states, that the maxima and minima of total ozone approximately coincide with the troughs and ridges of the 500 mb height field. Here we further examine the negative correlation between total ozone and middle tropospheric pressure.

Fig. 18 shows the January-mean total ozone and 500 mb height at 50°N latitude. The upper panel shows the deviations of the total ozone and 500 mb height from their respective zonal means, with the total ozone plotted on an inverted scale. Here $\hat{\Omega}^* = \hat{\Omega} - [\hat{\Omega}]$ and $\hat{Z}^*_{500} = \hat{Z}_{500} - [\hat{Z}_{500}]$, where Ω is the total ozone, Z_{500} is the 500 mb geopotential height, the overbar denotes the time mean, the bracket denotes the zonal mean, and the asterisk denotes the deviation from the zonal mean. The correlation between $\hat{\Omega}^*$ and $\hat{Z}^*_{500}$ is -0.724. These curves, as well as the scatter diagram in the lower panel of the figure, show that the agreement between $\hat{\Omega}$ and $\hat{Z}_{500}$ is much better in the Eastern Hemisphere than in the Western Hemisphere.

The top panel of Fig. 19 shows the relation between total ozone and 500 mb geopotential height at 50°N latitude at 0600 GMT 23 January (cf. middle panel of Fig. 14), again with the scale for total ozone inverted. The correlation between Ω^* and Z^*_{500} is -0.821, and the agreement is again better in the Eastern Hemisphere than in the Western Hemisphere.

In the central panel of Fig. 19 we show the transient eddy components, $\Omega'^* = \Omega^* - \hat{\Omega}^*$ and $Z'^*_{500} = Z^*_{500} - \hat{Z}^*_{500}$, where the prime denotes the deviation from the time mean. The correlation between Ω'^* and Z'^*_{500} is -0.870, and the agreement in the Western Hemisphere is as good as in the Eastern Hemisphere. The rms height deviation from the linear regression line is 34.0 m. This is about half of the contour interval that is routinely used to depict the 500 mb height field (as in Fig. 3).

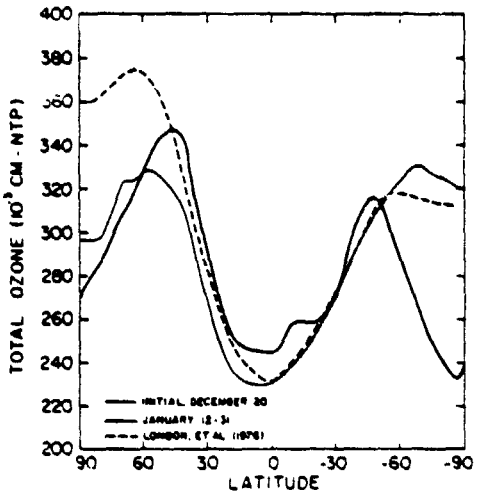


Fig. 17. Simulated and observed latitudinal distributions of total ozone.

b. *Vertical distribution*

Fig. 20 shows meridional cross sections of the zonally averaged initial O_3 mixing ratio, the simulated January distribution, and the observed distribution for December–February.

During the simulation the ozone maximum at 10 mb became more intense and its center shifted from the winter to the summer side of the equator, near 10°S. Although this is not the location of the observed maximum shown in the bottom panel of Fig. 20, both Dütsch (1969) and Kreuger *et al.* (1973) show the maximum on the summer side of the equator. In general, above the 10 mb level the simulated distribution shows good agreement with the observed field given by Krueger *et al.* (1973). The principal shortcoming of the simulation is that between about 25 and 200 mb the isopleths slope upward instead of downward from the middle latitudes toward the poles in both hemispheres. This is why the simulated total ozone is less than the observed total ozone in the polar regions, as was shown in Fig. 17. Within the troposphere, the simulated zonally averaged O_3 mixing ratio shows a maximum in each hemisphere, with the axes of the maxima sloping downward toward the tropics. These maxima, which are also present in the observed ozone distribution, are located near the subtropical descending branches of the mean-meridional mass circulation, shown in Fig. 9. However, the simulated O_3 maximum in the Northern Hemisphere troposphere is located about 20° farther from the pole than is the observed maximum.

5. Simulated ozone sources, sinks and transports

In the following, we show the simulated O_3 sources, sinks and transports, and discuss their relative roles in the atmospheric O_3 budget. The form

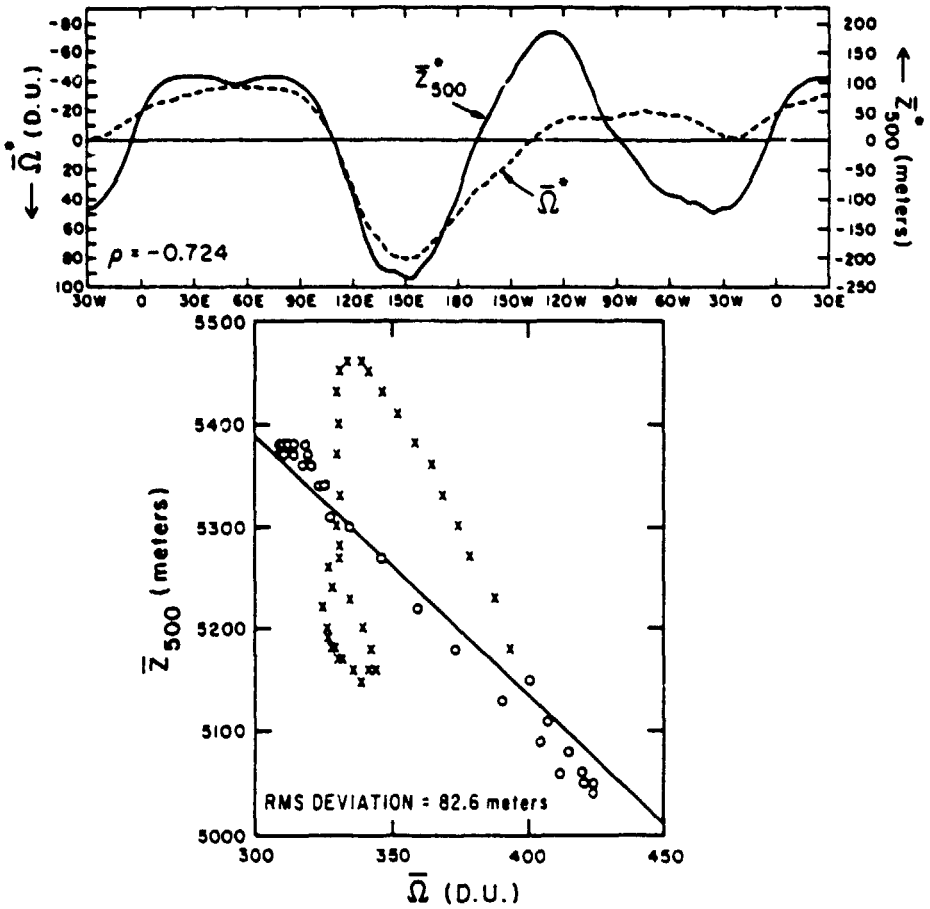


FIG. 18. Top: Deviation of total ozone and 500 mb geopotential height from their respective zonal means at 50°N, averaged for the last 20 days of the integration (12–31 January). Bottom: Scatter diagram of total ozone and 500 geopotential height at 50°N, averaged for the last 20 days of the integration (12–31 January). Western Hemisphere (crosses); Eastern Hemisphere (circles).

in which the results are presented is described in the Appendix.

a. Sources and sinks

1) TIME-DEPENDENT BEHAVIOR

Fig. 21 shows the globally integrated photochemical production rate of ozone, the globally integrated surface destruction rate of ozone and the global-mean total ozone as functions of time. We see that, immediately following the ozone re-initialization on 20 December, the photochemical production rate increased by an order of magnitude. It stayed high for about a day and then gradually decreased toward a negative value near 10 January. During the remainder of the simulation the production rate slowly increased.

In addition to the long-term variation, the production rate shows variations which have periods of a day and less. To our knowledge this high-frequency variation in the global photochemical O_3 production rate was not found in any of the other simulations that have been made, whether with purely photochemical models or with dynamical-photochemical models. Oscillations in the globally integrated photochemical O_3 production rate with periods of a day and less can occur only if there are longitudinal variations in the O_3 distribution. If the ozone did not vary with longitude, the global production rate distribution would simply travel around the earth with the sun, and the globally integrated production rate would be constant in time. As was shown in Fig. 16, the time-averaged global distribution of total ozone varies in the longitudinal direction as a consequence of the atmospheric circulation, and it is this longitudinal variation, interacting with the photochemical production, which produces the diurnal and higher frequency variations in the total global photochemical production rate.

The globally integrated surface destruction rate is shown in Fig. 21, with a scale that is 50 times larger

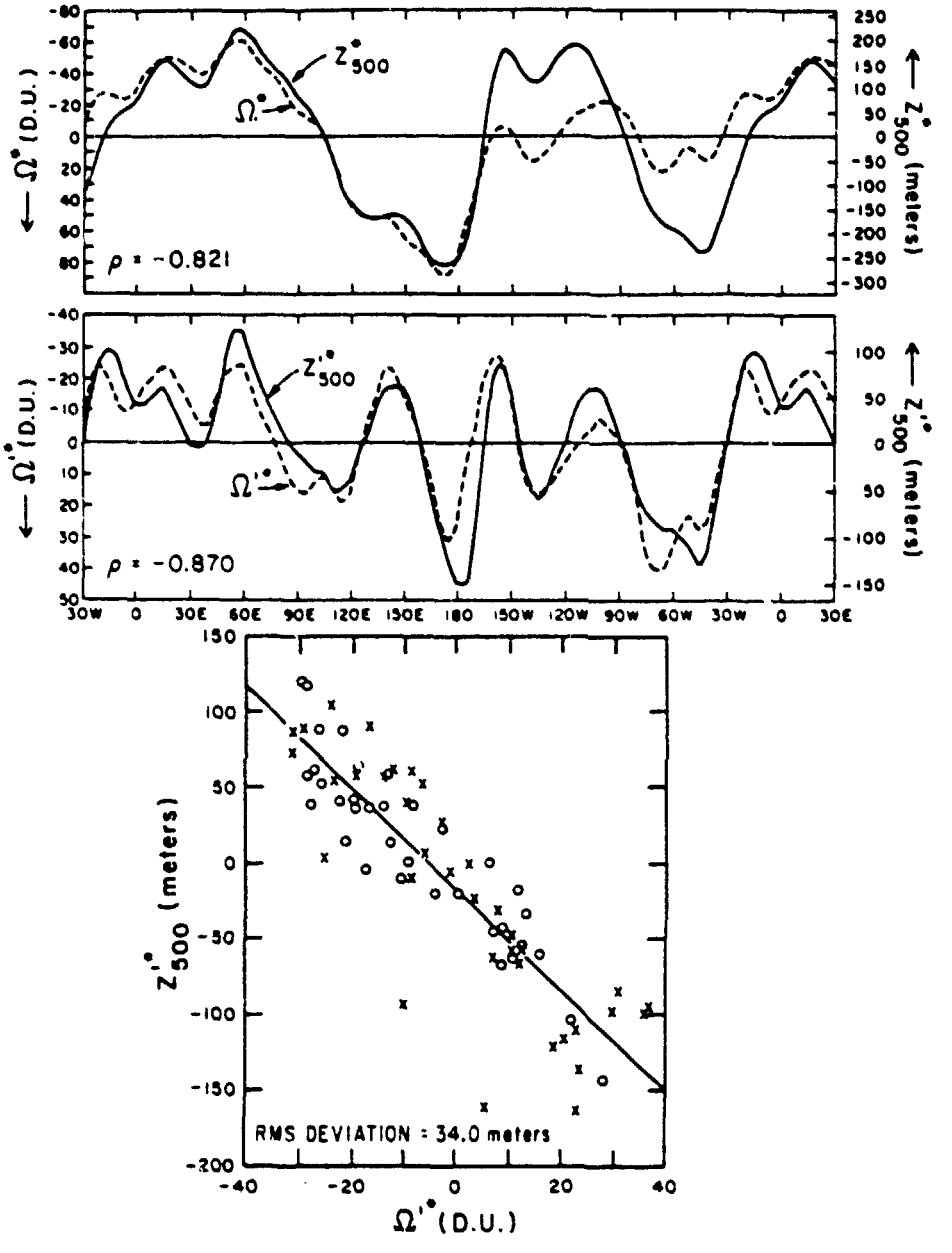


FIG. 19 Top: Deviation of total ozone and 500 mb geopotential height from their respective zonal means at 50°N, at 0600 GMT 23 January. Middle: Deviation of total ozone and 500 mb geopotential height from their respective zonal and January time means at 50°N, at 0600 GMT 23 January. Bottom: Scatter diagram of the data in the middle panel. Western Hemisphere (crosses); Eastern Hemisphere (circles).

than the scale for the photochemical production rate. We see that the surface destruction rate decreased slightly from the time of O_3 re-initialization to 23 December, remained nearly constant for a few days, and then increased. At the end of the integration period, the surface destruction rate is within the range of the estimates based on observations (CIAP, 1974).

The global-mean total ozone as a function of time is shown in the bottom panel of Fig. 21. The high-frequency oscillations, which are so evident in the photochemical production rate curve, are not seen in the curve of global-mean total ozone because, taking 30 tons s^{-1} as the mean amplitude of the diurnal production rate oscillation, only about 10^6 tons of ozone are created or destroyed during each

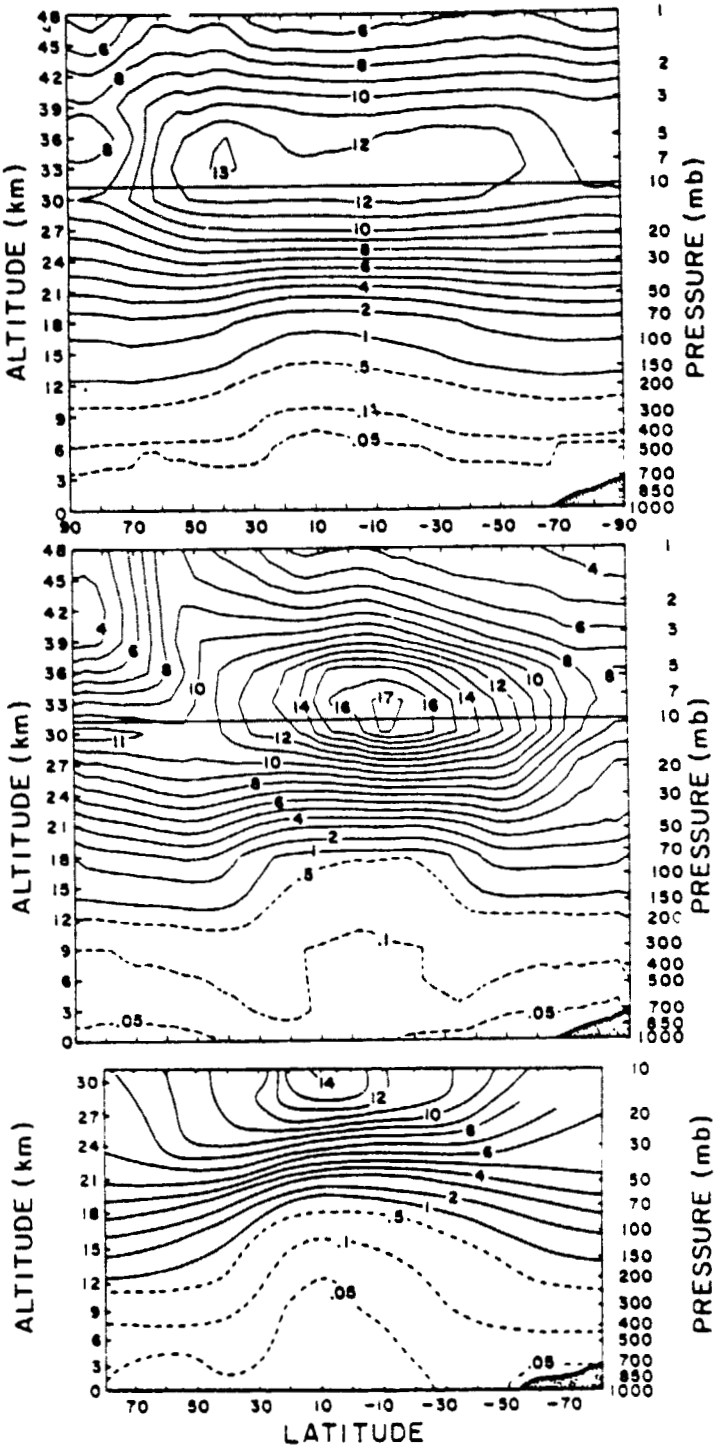


FIG. 20. Zonally averaged ozone mixing ratio ($\mu g\ g^{-1}$). Top: Initial ozone on 20 December. Middle: Simulated ozone, averaged for the last 20 days of the integration (12–31 January). Bottom: Observed normal for December–February (from Newell *et al.*, 1974).

half of the diurnal cycle. This corresponds to a change in total ozone of only 0.09 D.U., which is too small to be seen on the scale of Fig. 21.

From the time of O_3 re-initialization to 30 December, the global-mean total ozone increased from 283.4 D.U. to a maximum of 291.7 D.U. be-

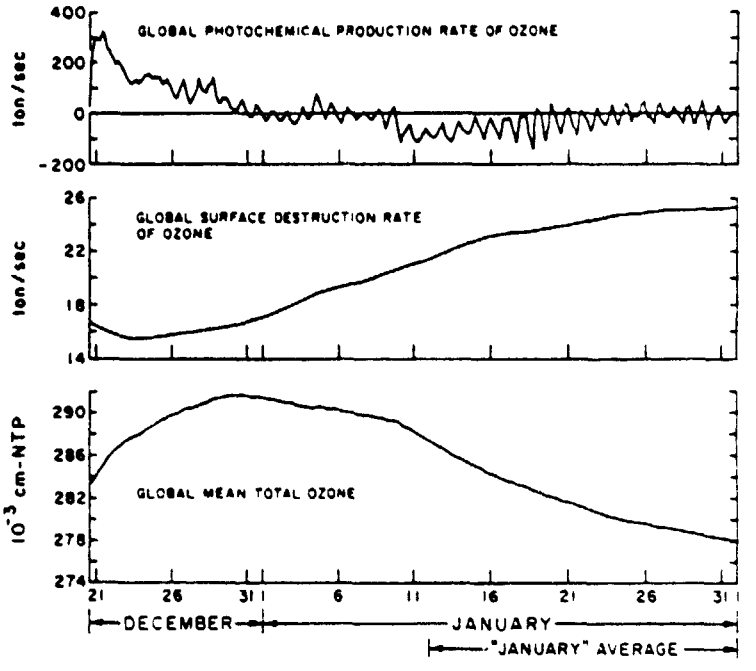


FIG. 21. Global photochemical production rate of ozone (top), the global surface destructure rate of ozone (middle), and the global-mean total ozone (bottom), as functions of time.

cause the photochemical production rate was larger than the surface destruction rate. After that, the total ozone decreased.

2) JANUARY MEAN

Fig. 22 shows the simulated zonally integrated January photochemical O_3 production rate. Above about 42 km, there is photochemical production of ozone in almost all latitudes, with the principal maximum near 10°S and a secondary maximum near 40°N. Between about 30 and 16 km, there is another large region of photochemical O_3 production, with the main maximum at about 25 km and 20°S. A smaller maximum is located at about 30 km and 15°N in an O_3 production region which extends into the upper stratosphere. (No photochemical production of ozone was calculated for the model layers below 100 mb.) There is a region of large photochemical destruction of ozone in the tropics, between about 41 and 30 km, with its maximum at about 36 km and 10°S, and another region of large O_3 destruction rate in the Northern Hemisphere, north of about 20° latitude, with its maximum at about 28 km and 40°N. There is no photochemical production or destruction of ozone in the latitudes of the polar night.

Fig. 23 shows the latitudinal distribution of the zonally and vertically integrated photochemical O_3 production rate, together with the zonally integrated surface O_3 destruction rate. We see that for the atmospheric column which extends from the surface to the 1 mb level, the photochemical production is a source of ozone in the tropics and subtropics of the Southern Hemisphere and in the tropics of the Northern Hemisphere, and is a sink everywhere else (except in the polar-night latitudes). The ozone is destroyed at the earth's surface everywhere, but at a maximum rate in the tropics. We shall discuss the distributions of the O_3 production and destruction rates that are shown in Figs. 22 and 23 after we have presented the O_3 transports.

b. Transports

In this section we present the transports of ozone by the atmospheric circulation: the time-averaged zonally integrated meridional O_3 transport, the time-averaged zonally integrated vertical O_3 transport and the convergence of these transports. As described in the Introduction, two different transport processes have been proposed to explain the discrepancies between the observed latitudinal and seasonal O_3 distribution and that which is predicted from photochemical theory. One is the O_3 transport by the mean-meridional mass circulation, and the other is the O_3 transport by the zonal atmospheric waves or eddies. To determine how these two transport processes operate in the model simulation, we separate the total transports and the total transport convergences into mean-meridional and zonal-eddy components (see Appendix).

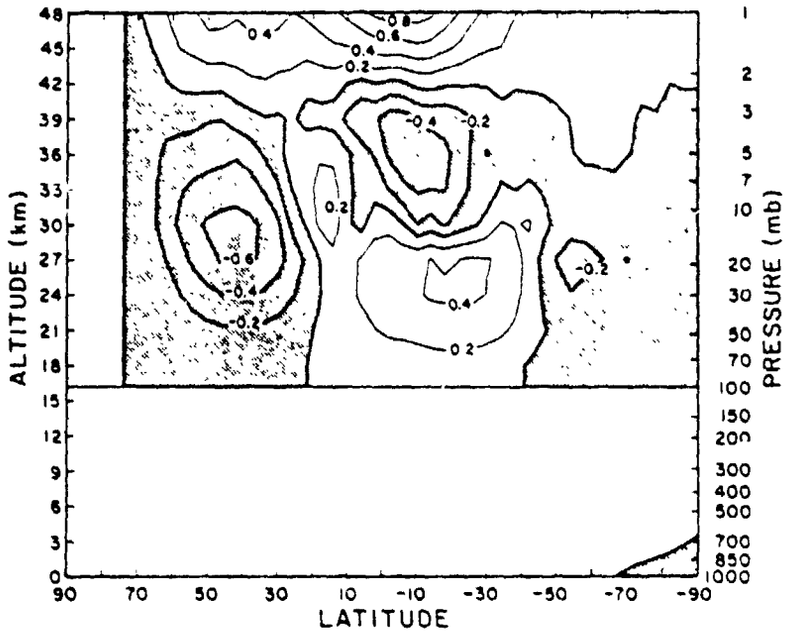


FIG. 22. Zonally integrated simulated photochemical ozone production rate [tons s^{-1} per 4° latitude per km vertical height], averaged for the last 20 days of the integration (12–31 January). Photochemical ozone destruction is shaded.

1) MERIDIONAL (NORTH–SOUTH) TRANSPORT

Fig. 24 shows the simulated January meridional O_3 transports by the mean-meridional circulation and the zonal eddies, and the sum of the two transports. The meridional O_3 transport by the mean-meridional circulation, shown in the top panel, agrees in sign with the meridional mass transport by the mean-meridional circulation shown in Fig. 9. In the tropics, the Northern (winter) Hemisphere Hadley cell, which crosses the equator, transports ozone southward in the lower part of the troposphere and northward in the upper troposphere and stratosphere. The Southern (summer) Hemisphere Hadley cell is narrower and more shallow, and transports ozone northward below about 6 km and southward between 6 and 20 km. In the Northern Hemisphere middle latitudes, the deep Ferrel cell transports ozone toward the pole in the lower troposphere, and toward the equator in the middle and upper troposphere and over the entire depth of the stratosphere. Corresponding transports are produced by the Southern Hemisphere Ferrel cell and by the higher latitude circulation cells in both hemispheres.

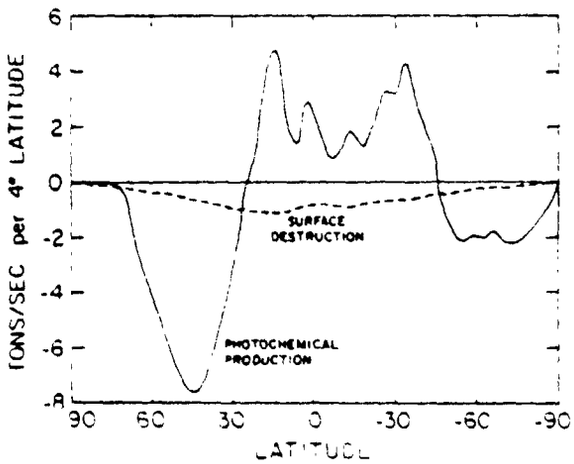


FIG. 23. Zonally and vertically integrated photochemical O_3 production rate and the zonally integrated surface destruction rate [tons s^{-1} per 4° latitude], averaged for the last 20 days of the integration (12–31 January).

The meridional O_3 transport by the zonal eddies is shown in the middle panel of Fig. 24. In the troposphere and lower stratosphere of both hemispheres, the tropical eddy transport is toward the equator, the middle-latitude eddy transport is toward the poles, and the high-latitude eddy transport is toward the equator. In general, the eddy transports of ozone are larger in the Northern (winter) Hemisphere.

Comparing the meridional O_3 transports by the zonal eddies with the transports by the mean-meridional circulation, we see that they are of the same order of magnitude and that, above about 6 km, the two transports are generally in the opposite directions. However, as shown by the sum of the two transports in the bottom panel of Fig. 24, the transports by the mean-meridional circulation and by the atmospheric eddies do not cancel one another. One or the other transport process dominates, and the

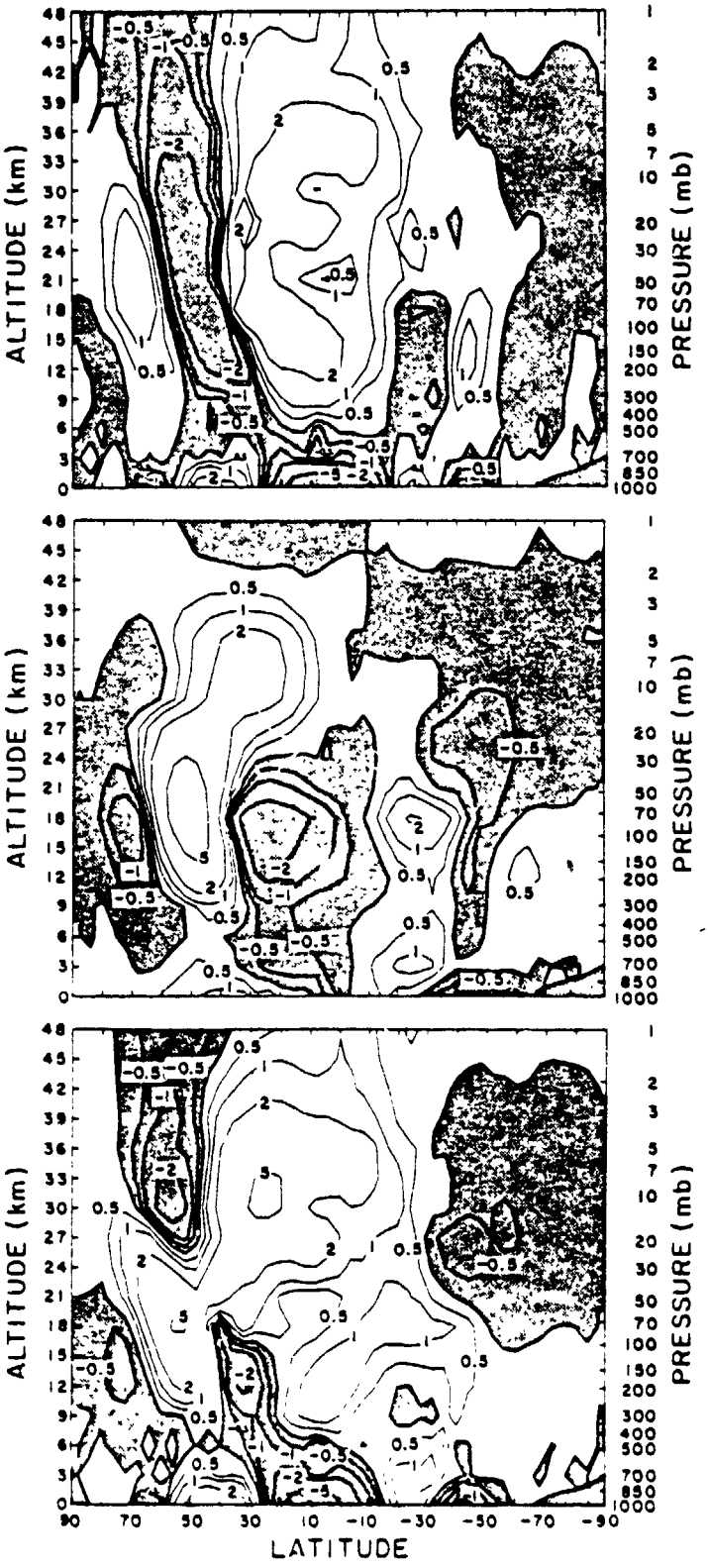


FIG. 24. Meridional transports of ozone [tons s^{-1} per km vertical height], averaged for the last 20 days of the integration (12–31 January). Top: Transport by the mean-meridional circulation. Middle: Transport by the zonal eddies. Bottom: Total meridional transport. Southward transport is shaded.

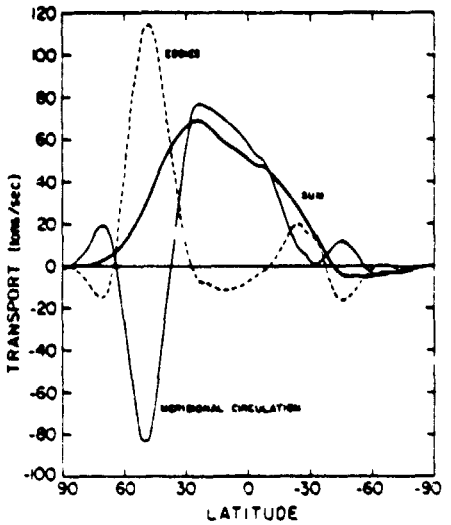


FIG. 25. Vertically integrated meridional transports of ozone, averaged for the last 20 days of the integration (12–31 January).

magnitude of the net transport is comparable to that of the two components. In general, the total meridional O_3 transport has larger magnitudes in the Northern (winter) Hemisphere.

Fig. 25 shows the vertically integrated meridional transports of ozone by the mean-meridional circulation, by the zonal eddies and their sum. The transport by the mean-meridional circulation dominates in the tropics and carries ozone from the Southern (summer) Hemisphere into the Northern (winter) Hemisphere. In the tropics of the Northern Hemisphere, the eddies produce only a weak vertically integrated southward transport, which is the small difference between a large northward transport above about the 25 km level and large southward transport below that level, as was shown in Fig. 24. In the middle latitudes of both hemispheres, the vertically integrated mean-meridional transport of ozone is away from the poles and the eddy transport is toward the poles. The two transports are of comparable magnitude in the middle latitudes and have their maxima at about 50°N and 45°S; but the eddy transport dominates, so that in the middle latitudes the direction of the total meridional O_3 transport is in the direction of the eddy transport. Poleward of about 60°N latitude, both the vertically integrated mean-meridional transport and eddy transport have changed sign, but here it is the mean-meridional transport which dominates and, therefore, here also the net transport is toward the poles.

2) VERTICAL TRANSPORT

Fig. 26 shows the simulated January vertical O_3 transports by the mean-meridional circulation and the zonal eddies, and the sum of the two transports. The vertical transport by the mean-meridional circulation, shown in the top panel of Fig. 26, agrees in sign with the mass transport by the mean-meridional circulation, shown in Fig. 9. Within the troposphere, the descending branches of the circulation cells carry more ozone than the ascending branches.

The vertical transport of ozone by the zonal eddies, shown in the middle panel of Fig. 26, is downward almost everywhere in the troposphere and lower stratosphere, and upward over most of the middle and upper stratosphere. This shows that, with respect to the time and zonally averaged O_3 mixing ratio (shown in the middle panel of Fig. 20), the vertical eddy O_3 transport is a downgradient transport over most of the atmosphere, but that from about 30 km down to as low as 10 km, the vertical eddy transport of ozone is a countergradient transport.

Comparing the vertical eddy transport (the middle panel of Fig. 26) with the meridional eddy transport (the middle panel of Fig. 24) shows that in the middle latitudes of the lower stratosphere the eddies transport ozone downward and poleward. This is what we expect from the known wave motions in the lower stratosphere, where the poleward-moving warm (and ozone-rich) air sinks and the equatorward-moving cold (and ozone-poor) air rises: In the troposphere, between about 35°N and 40°S, the eddies transport ozone downward and toward the equator. This is what we expect from the known wave motions in the troposphere, where the poleward-moving warm (and ozone-poor) air rises and the equatorward-moving cold (and ozone-rich) air sinks. However, in the middle latitudes, especially in the Northern Hemisphere, the eddies transport ozone downward and poleward. This is a change from the earlier results presented by Mintz and Schlesinger (1975) which showed that the eddy transport was downward and equatorward in the tropospheric middle latitudes. Why this feature of the 20-day average eddy transport differs from what we expect from baroclinic eddy theory is not understood and is being investigated using the simulation data.

As shown in the top and middle panels of Fig. 26, the vertical transport of ozone by the mean-meridional circulation has a structure which varies mainly in the meridional direction, and the vertical transport of O_3 by the eddies has a structure which varies mainly in the vertical direction. This means that the two vertical transport processes will alternately be in phase and out of phase in the north-south direction. But the vertical O_3 transport by the mean-meridional circulation is almost everywhere larger in magnitude than the vertical transport by the eddies. Consequently, the total vertical O_3 transport, shown in the bottom panel of Fig. 26, is very similar in pattern to the vertical O_3 transport by the mean-meridional circulation itself. In the middle latitudes of both hemispheres, below about 27

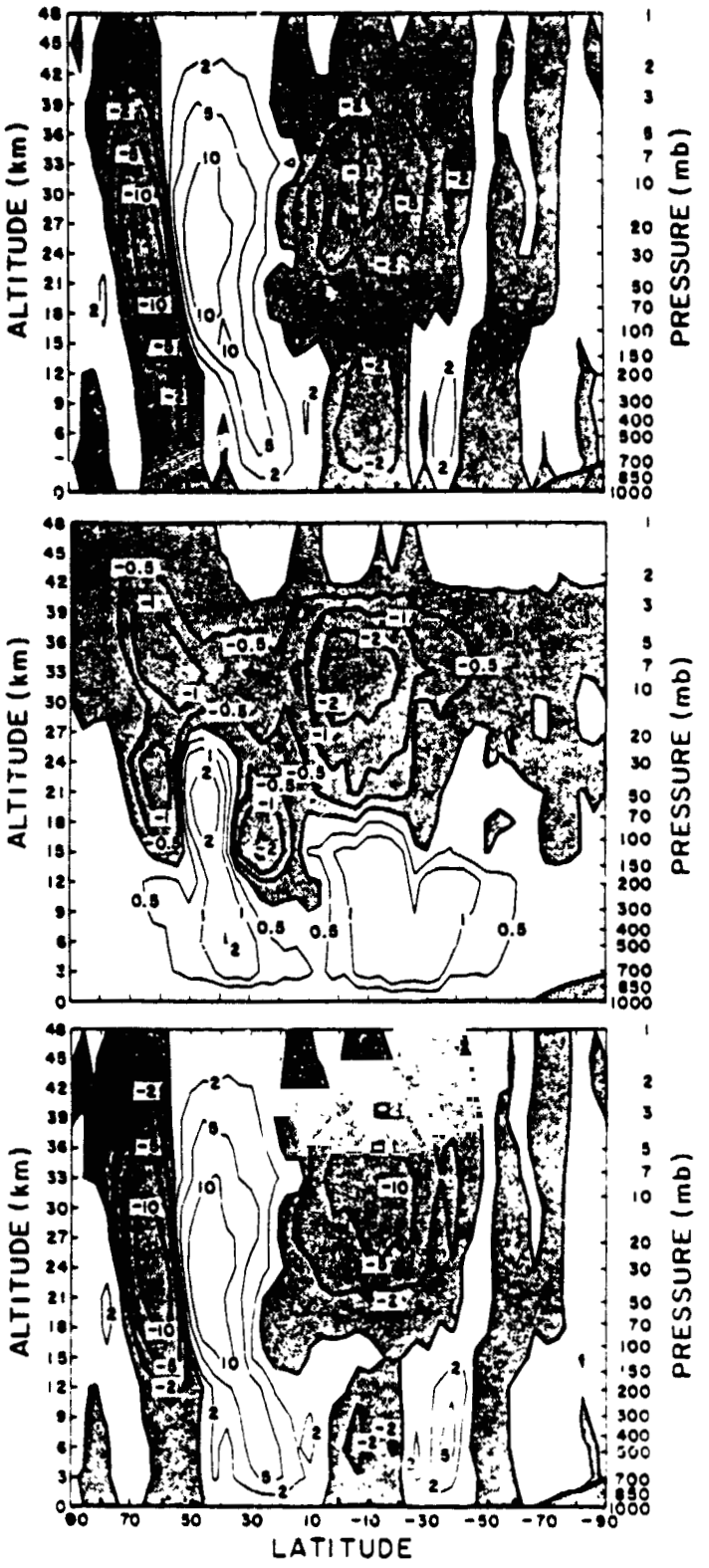


FIG. 26. Vertical transports of ozone [tons s^{-1} per 4° latitude], averaged for the last 20 days of the integration (12–31 January). Top: Transport by the mean-meridional circulation. Middle: Transport by the zonal eddies. Bottom: Total vertical transport. Upward transport is shaded.

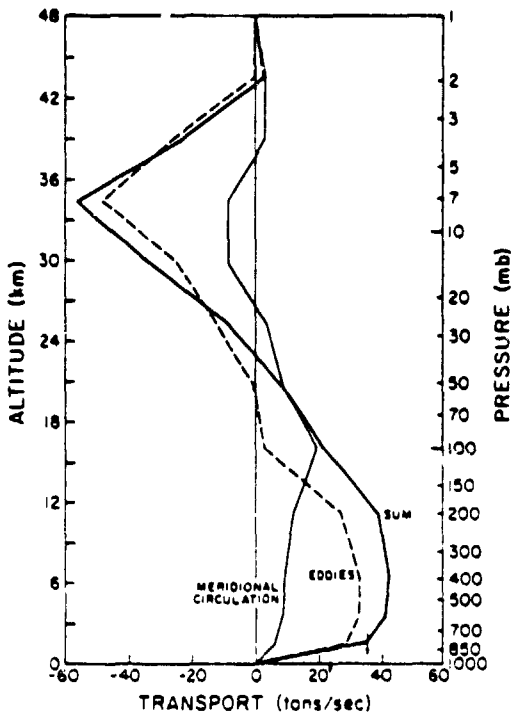


FIG. 27. Zonally and latitudinally integrated vertical transports of ozone, averaged for the last 20 days of the integration (12–31 January). Positive values are downward ozone transports. The solid arrow shows the downward ozone transport into the lowest model layer, the dashed arrow shows the surface destruction rate.

km, the vertical transport by the eddies is in phase with the transport by the mean-meridional circulation, and this produces the pronounced downward total transport, with maxima that slope downward toward the equator. It is this cooperative downward transport which causes the O_3 mixing ratio isopleths to dip downward toward the equator in the middle latitudes of the troposphere (as seen in the middle panel of Fig. 20), and which supports the surface destruction rate maxima in the tropics (shown in Fig. 23). This is in contrast to the upward transport by the mean-meridional circulation, just south of the equator, which makes the lines of constant O_3 mixing ratio bulge upward relative to their elevation in higher latitudes.

Fig. 27 shows the zonally and latitudinally integrated vertical O_3 transports, as obtained from the pole-to-pole integration of the transports shown in Fig. 26. The transport by the mean-meridional circulation is downward in the troposphere and lower stratosphere, upward in the middle stratosphere, and almost zero in the upper stratosphere. The transport by the eddies has a similar distribution. Except near the tropopause, the vertical transport by the eddies is several times larger than the transport by the mean-meridional circulation. This dominance by the eddies is due to the fact that the vertical transport by the mean-meridional circulation changes sign with latitude, whereas the vertical transport by the eddies changes sign with altitude, as was shown in Fig. 26. Thus it is that in the global (i.e., one-dimensional) sense, the eddies are more important than the mean-meridional circulation in transporting ozone from one level of the atmosphere to another. In this one-dimensional sense, moreover, the upward eddy transport of O_3 between 20 and 31 km is a countergradient transport.

3) TRANSPORT CONVERGENCES

Fig. 28 shows the meridional distribution of the convergence of the O_3 transport as simulated for January. The top panel shows the convergence of the O_3 transport by the mean-meridional circulation (the convergence of the meridional transport by the mean-meridional circulation shown in the top panel of Fig. 24, plus the convergence of the vertical transport by the mean-meridional circulation shown in the top panel of Fig. 26). The middle panel shows the convergence of the O_3 transport by the zonal eddies (the convergence of the meridional eddy transport shown in the middle panel of Fig. 24, plus the convergence of the vertical eddy transport shown in the middle panel of Fig. 26). The bottom panel shows the sum of all the O_3 transport convergences.

We see that the convergences of the transports by the mean-meridional circulation and by the eddies are in large measure out of phase. They partially cancel one another, but a well-organized residual field of somewhat smaller magnitude remains, as seen in the bottom panel of Fig. 28.

If we compare the bottom panel of Fig. 28 with Fig. 22 and the dashed curve in Fig. 23, we see that in a general way there is a one-to-one correspondence between the O_3 transport divergence (convergence) and the O_3 sources (sinks). The two large photochemical sinks, one centered at 40°N and 28 km elevation and the other centered at 10°S and 36 km elevation, are in regions of large O_3 transport convergence; the large photochemical source, centered at 20°S and 25 km elevation, is in a region of large O_3 transport divergence; and the large surface sink, with its maximum in the tropics, agrees with the transport convergence within the lowest model layer.

The comparison of Fig. 22 and the bottom panel of Fig. 28 enables us to draw some conclusions about the influence of the O_3 transports on the O_3 production. For example, the region of large O_3 production, which is centered at 20°S and 25 km elevation, is one in which the divergence of the transport, due mainly to the ascending branch of the Hadley circulation and the countergradient vertical eddy transport, maintains the O_3 concentration at less than its photochemical equilibrium value; whereas, immediately above this level, these vertical trans-

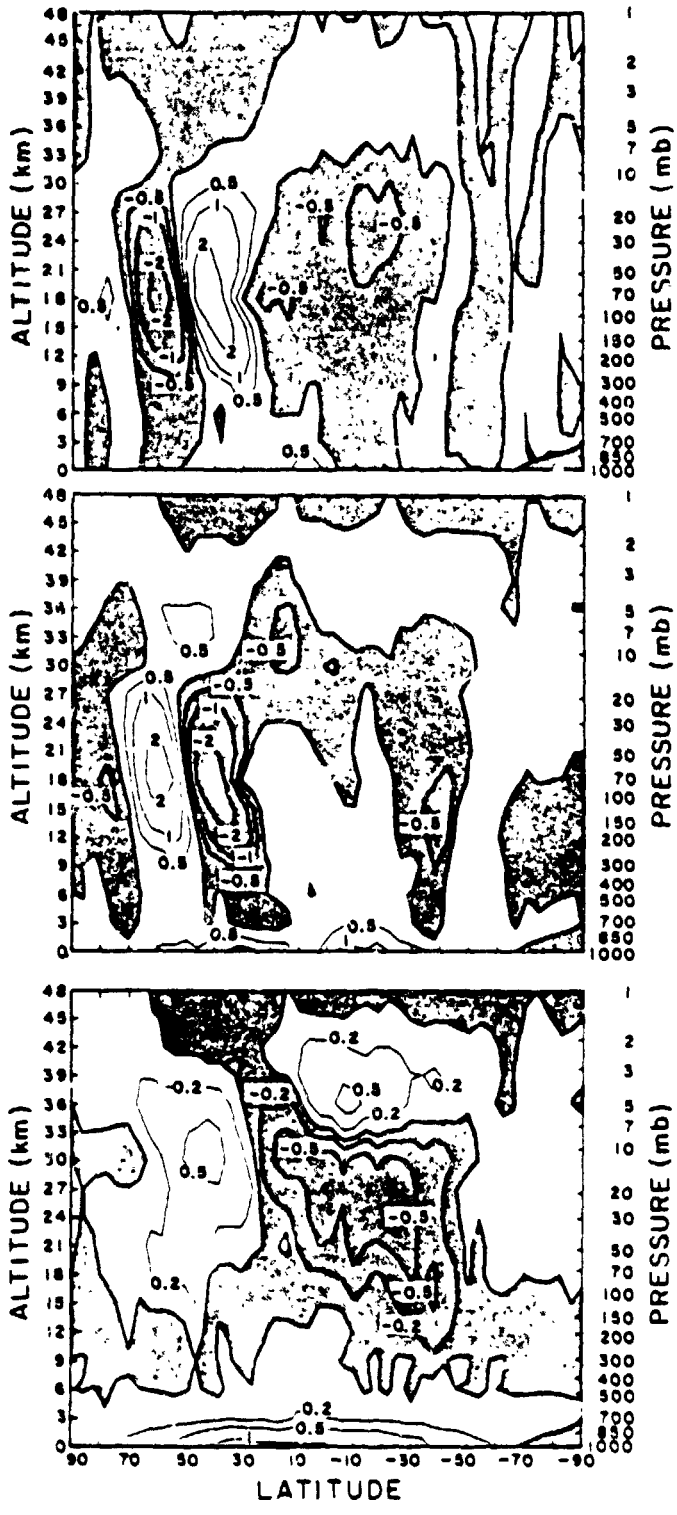


FIG. 28. Convergence of the O_3 transport [tons s^{-1} per 4° latitude per km vertical height], averaged for the last 20 days of the integration (12–31 January). Top: Convergence of O_3 transport by the mean-meridional circulation. Middle: Convergence of O_3 transport by the zonal eddies. Bottom: Convergence of the total O_3 transport. Regions of transport divergence are shaded.

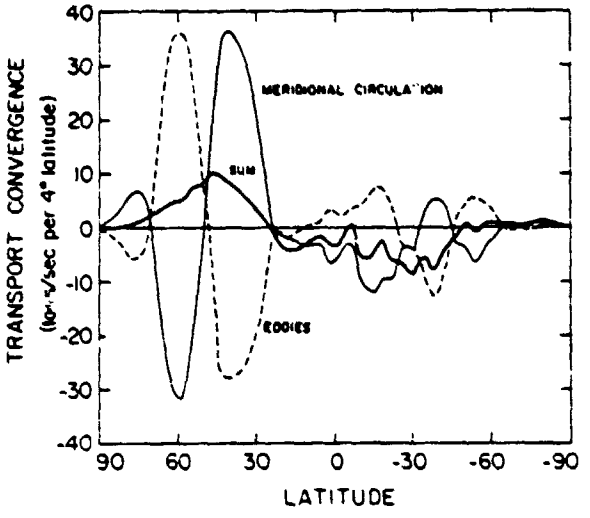


FIG. 29. Vertical integrals of the convergence of the ozone transport by the mean-meridional circulation and by the zonal eddies, averaged for the last 20 days of the integration (12–31 January).

port processes maintain the ozone concentration at a value greater than the photochemical equilibrium value. Similarly, in the middle and higher latitudes, it is the poleward transports by both the eddies and the mean-meridional circulation which maintain the ozone concentration at a value greater than its photochemical equilibrium value.

Fig. 29 shows the vertical integrals of the convergences of the ozone transports by the mean-meridional circulation and by the zonal eddies, and the sum of the two vertically integrated convergences. These curves correspond, of course, to the derivatives of the poleward transports shown in Fig. 25. Again, we see the high degree of compensation between the two processes. The heavy curve in Fig. 29 is shown again in Fig. 30, together with the vertically integrated photochemical O_3 production rate and the surface destruction rate (taken from Fig. 23) and the sum of these terms. As expected, in the tropics and middle latitudes of the Southern Hemisphere, ozone is being decreased by meridional transport divergence and increased by the photochemistry. Poleward of this region in both hemispheres, ozone is being increased by the meridional transport convergence and destroyed by the photochemistry. In most latitudes the magnitude of the surface destruction is much smaller than the magnitudes of the transport convergence and photochemical production.

6. Summary and discussion

The photochemical production and destruction of ozone in the atmosphere is greatly dependent on the atmospheric motion field. In some regions, the O_3 transport divergence maintains the local O_3 concentration below its photochemical equilibrium value and this leads to photochemical O_3 production. In other regions, the O_3 transport convergence maintains the local O_3 concentration above its photochemical equilibrium value and this leads to photochemical O_3 destruction.

The motion field, in turn, depends on the atmospheric heating field and, in the stratosphere, the heating depends in large part on the O_3 absorption of solar radiation. But the motions in the stratosphere are not governed solely by the radiational heating of the stratosphere. To a large extent the stratospheric circulation is driven by a mechanical transfer of energy from the underlying troposphere. Thus it is that the distribution of ozone and its photochemical sources and sinks depend on the general circulation of the troposphere and stratosphere as a coupled system.

Inasmuch as the O_3 transports and transport divergences in large measure depend on the vertical air velocities, and these vertical velocities cannot be directly determined from observations with the required accuracy and distribution in space and time, we must resort to the method of numerical simulation with an atmospheric general circulation model to learn how the global ozone is produced and transported.

In the Introduction the question was posed: Is the net poleward ozone transport produced by the mean-meridional circulation or by the zonal eddies in the atmospheric motion field? The answer is—they both contribute. In tropical and subtropical latitudes, and within the polar caps, the mean-meridional cir-

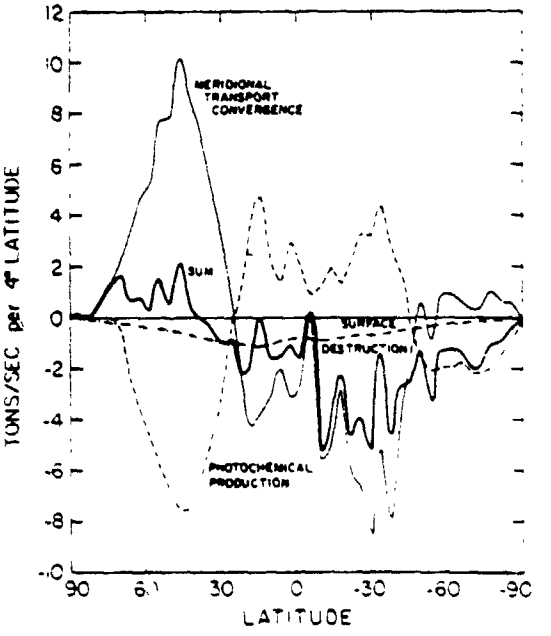


FIG. 30. Vertically-integrated photochemical O_3 production rate, surface destruction rate, convergence of the O_3 transport, and the sum of these terms, averaged for the last 20 days of the integration (12–31 January).

culation transports ozone poleward and dominates; in the middle latitudes, the zonal eddies transport ozone poleward and dominate. This is in agreement with other general circulation model transport calculations such as that of Hunt (1969).

We have found, furthermore, that the net vertical O_3 transport (the transport averaged over all longitudes and latitudes) is dominated by the zonal eddies, and that over the atmospheric layer between 20 and 31 km elevation, this transport is a countergradient transport. This result, however, may be due to the fact that the simulated ozone did not reach equilibrium or to the simplicity of the model's photochemistry.

We have seen that the numerical simulation produced a large negative correlation between total ozone and middle-troposphere pressure, especially with respect to the transient waves. This correlation has long been known from observations, but the reason for it has remained obscure. But, in the simulation, all of the terms in the ozone budget and all the components of the atmospheric motion, mass and heating fields are recorded with high precision and resolution in space and time, so that from these data it should be possible to find the reason for the high negative correlation. In fact, Schlesinger is undertaking that task with the data set generated in this simulation experiment.

There is a possible practical application of the high negative correlation between total ozone and pressure in the middle troposphere, namely, the determination of a reference height for numerical weather prediction initialization in data sparse regions. If we can obtain an accurate measure of the total ozone by remote sensing from satellites, then, as we have seen, that measurement might be used to determine the 500 mb height to something like 35 m rms error.

On the other hand, if we wish to determine the three-dimensional distribution of ozone from incomplete ozone observations, the best way may be to use an atmospheric general circulation model and the technique of four-dimensional data assimilation as, for example, was done by Miyakoda *et al.* (1976) to determine the three-dimensional distribution of the wind field from incomplete wind observations. In this technique, continuous insertion of incomplete observational data into an atmospheric general circulation model which carries the dependent variables forward in time, will use information that has a good resolution in time as a substitute for good resolution in space. The three-dimensional, synoptic ozone distributions which will be obtained by such a dynamical analysis of the ozone observations should be superior to the conventional static forms of analysis of the ozone observations.

Acknowledgments. We wish to thank especially Professors A. Arakawa and S. V. Venkateswaran who participated in the initial formulation of the research plan for this study and who closely followed and assisted in its execution. We also thank the other members of the research group who contributed to the design and construction of the UCLA atmospheric general circulation model: Drs. A. Katayama, J.-W. Kim, S. Lord, D. A. Randall, W. H. Schubert, T. Tokioka and W. Chao. And we thank Mrs. D. Hollingsworth and Mr. C. Kurash for their programming contributions.

We express our appreciation to Mr. W. B. Kehl, Director of the UCLA Campus Computing Network; Professor W. J. Dixon, Director of the UCLA Health Sciences Computing Facility; and Dr. M. Halem of the NASA Goddard Institute for Space Studies, for providing the computing time for this study.

We thank Dr. J. D. Mahlman of the NOAA Geophysical Fluid Dynamics Laboratory, Princeton, for his constructive review of the manuscript.

The study was supported by the Global Atmospheric Research Program, Division of Atmospheric Sciences, National Science Foundation, under Grant GA-34308; by the U. S. Department of Transportation, Climatic Impact Assessment Program, under Grant GA-34306X; and by the National Aeronautics and Space Administration, Goddard Space Flight Center, Institute for Space Studies, under Grant NGR 05-007-328.

APPENDIX

Ozone Continuity Equations

In spherical and σ-coordinates the ozone conservation equation may be written in flux form as

$$\frac{\partial}{\partial t}(a^2 \cos\phi\pi q) + \frac{\partial}{\partial \lambda}(a\pi uq) + \frac{\partial}{\partial \phi}(a\cos\phi\pi vq) + \frac{\partial}{\partial \sigma}(a^2\cos\phi\pi\dot{\sigma}q) = a^2\cos\phi\pi\dot{O}, \quad \text{(A1)}$$

where t, λ, ϕ are time, longitude and latitude; σ is the vertical coordinate defined by

$$\sigma = (p - p_t)/\pi, \quad \text{(A2)}$$

$$\pi = \begin{cases} \pi_u = p_t - p_T, & p_T \le p < p_t \\ \pi_L = p_S - p_t, & p_t < p \le p_S, \end{cases} \quad \text{(A3)}$$

with $p_T = 1$ mb, $p_t = 100$ mb and p_S the surface pressure, which varies with λ, ϕ and t; q is the ozone mixing ratio; u, v, $\dot{\sigma}$ are the zonal, meridional and vertical velocity components; a is the earth's radius; and $\dot{O}$ is the ozone source (and sink) per unit mass of air. Letting k denote the vertical index in the model ($0 \le k \le 24$), with the even-numbered k denoting the interfaces of the 12 model layers, the vertical finite-difference analog of (A1) used by the model is

$$\frac{\partial}{\partial t}(a^2\cos\phi\pi_k q_k) + \frac{\partial}{\partial\lambda}(a\pi_k u_k q_k) + \frac{\partial}{\partial\phi}(a\cos\phi\pi_k v_k q_k) + \frac{a^2\cos\phi}{\Delta\sigma_k}[(\pi\dot\sigma)_{k+1}\hat q_{k+1} - (\pi\dot\sigma)_{k-1}\hat q_{k-1}] = a^2\cos\phi\pi_k\dot O_k, \quad k \text{ odd}, \quad \text{(A4)}$$

where

$$\Delta\sigma_k = \sigma_{k+1} - \sigma_{k-1},$$

$$\pi_k = \begin{cases}\pi_u, & k < 14\\ \pi_L, & k < 14\end{cases} \quad \text{(A5)}$$

and $\hat q$ is the even-level O_3 mixing ratio given by the logarithmic-conserving vertical advection scheme (see Schlesinger, 1976a). Introducing

$$\left.\begin{aligned} m_k &= \pi_k a^2\Delta\phi\Delta\lambda\cos\phi, \quad f_k = \pi_k u_k a\Delta\phi,\\ g_k &= \pi_k v_k a\Delta\lambda\cos\phi\\ s_{k+1} &= (\pi\dot\sigma)_{k+1}a^2\Delta\phi\Delta\lambda\cos\phi\end{aligned}\right\}, \quad \text{(A6)}$$

(with $\Delta\lambda = 5°$ and $\Delta\phi = 4°$), Eq. (A2) may be written as

$$\frac{\partial}{\partial t}(m_k q_k) + \frac{\partial}{\partial x}(f_k q_k) + \frac{\partial}{\partial y}(g_k q_k) + \frac{1}{\Delta\sigma_k}\times[s_{k+1}\hat q_{k+1} - s_{k-1}\hat q_{k-1}] = m_k\dot O_k, \quad \text{(A7)}$$

where $x = \lambda/\Delta\lambda$ and $y = \phi/\Delta\phi$. Integrating the time-averaged form of (A7) around a latitude circle gives

$$\overline{\frac{\partial}{\partial t}2\pi'[m_k q_k]} + \frac{\partial}{\partial y}2\pi'\overline{[g_k q_k]} + \frac{1}{\Delta\sigma_k}(2\pi'\overline{[s_{k+1}\hat q_{k+1}]} - 2\pi'\overline{[s_{k-1}\hat q_{k-1}]}) = 2\pi'\overline{[m_k\dot O_k]}, \quad \text{(A8)}$$

where $\pi' = 3.14159\ldots$, square brackets denote the zonal average and the overbars the time average.

Because the mass-weighting term π, for m, g and s in (A8), is discontinuous across the 100 mb level, as shown by (A3), (A5), and (A6), we normalize the results by multiplying (A8) by $\Delta\sigma_k/(10g'\Delta z_k)$, where Δz_k is the nominal layer thickness shown in Fig. 1 and g' is the acceleration of gravity. Thus (A8) may be written as

$$\overline{\frac{\partial}{\partial t}}2\pi'\overline{[M_k q_k]} = -\frac{\partial}{\partial y}2\pi'\overline{[G_k q_k]} - (2\pi'\overline{[S_{k+1}\hat q_{k+1}]} - 2\pi'\overline{[S_{k-1}\hat q_{k-1}]}) + 2\pi'\overline{[M_k\dot O_k]}, \quad \text{(A9)}$$

where

$$M_k = \frac{C\Delta\sigma_k}{\Delta z_k}m_k, \quad G_k = \frac{C\Delta\sigma_k}{\Delta z_k}g_k, \quad S_{k+1} = \frac{C}{\Delta z_k}s_{k+1}, \quad C = 1/(10g'). \quad \text{(A10)}$$

Eq. (A9) is the form in which the results are presented. It expresses the fact that the time-averaged rate of change of ozone (in metric tons per second) within a latitudinal ring which is 4° wide and 1 km deep) is due, respectively, to the convergence of the time-averaged meridional ozone transport through the vertical sides, the convergence of the time-averaged vertical ozone transport across the horizontal boundaries, and the time-averaged ozone production or destruction within the ring. For the lowest model layer, the last term in (A9) has an additional component which represents the time-averaged destruction of ozone at the earth's surface.

REFERENCES

Ackerman, M., and C. Muller, 1972: Stratospheric nitrogen dioxide from infrared spectra. *Nature*, 240, 300–301.

Arakawa, A., 1966: Computational design for long-term numerical integration of the equations of fluid motion: Two dimensional incompressible flow. Part 1. *J. Comput. Phys.*, 1, 119–143.

——, 1970: Numerical simulation of large-scale atmospheric motion. *Numerical Solution of Field Problems in Continuum Physics*, Vol. 2, *SIAM-AMS Proceedings*, G. Birkhoff and S. Varga, Eds. Amer. Math. Soc., 24–40.

——, and Y. Mintz, 1974: Workshop notes on the UCLA atmospheric general circulation model (24 March–4 April 1974). Department of Meteorology, University of California, Los Angeles, 404 pp.

——, and W. H. Schubert, 1974: Interaction of cumulus cloud ensemble with the large-scale environment, Part I. *J. Atmos. Sci.*, 31, 674–701.

——, and V. R. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods in Computational Physics*, Vol. 17, J. Chang, Ed., Academic Press, 337 pp.

Barnett, J. J., 1974: The mean meridional temperature behavior of the stratosphere from November 1970 to November 1971 derived from measurements by the selective chopper radiometer on Nimbus 4. *Quart. J. Roy. Meteor. Soc.*, 100, 505–530.

Bjerknes, J., 1975: *Selected Papers*, M. G. Wurtele, Ed. Western Periodicals, 606 pp.

Brewer, A. W., 1949: Evidence for a world circulation provided by the measurements of helium and water vapor distribution in the stratosphere. *Quart. J. Roy. Meteor Soc.*, 75, 351–363.

CIAP, 1974: *The Natural Stratosphere of 1974. CIAP Monogr. 1*, A. J. Grobecker, Ed., Dept. of Transportation Climatic Impact Assessment Program, Washington, DC, 1199 pp.*

Chapman, S., 1930a: A theory of upper atmospheric ozone. *Memo. Roy. Meteor. Soc.*, 3, 103–125.

——, 1930b: On the annual variation of upper-atmospheric ozone. *Phil. Mag.*, 10, 345–352.

——, 1930c: On ozone and atomic oxygen in the upper atmosphere. *Phil. Mag.*, 10, 369–383.

Clark, J. H. E., 1970: A quasi-geostrophic model of the winter stratospheric circulation. *Mon. Wea. Rev.*, 98, 443–446.

Crutcher, H. L., and J. M. Meserve, 1970: *Selected Level Heights, Temperatures and Dew Points for the Northern Hemisphere*. NAVAIR Publ. 50-1C-52, Washington, D. C. [Available from U. S. Govt. Printing Office.]

Cunnold, D., F. Alyea, N. Phillips and R. Prinn, 1975: A three-dimensional dynamical chemical model of atmospheric ozone. *J. Atmos. Sci.*, 32, 170–194.

Dickinson, R. E., 1973: Method of parameterization for infrared cooling between the altitudes of 30 and 70 kilometers. *J Geophys. Res.*, 78, 4451–4457

Dobson, G. M. B., 1956: Origin and distribution of the poly-

atomic molecules in the atmosphere. *Proc. Roy. Soc., London*, **A236**, 187–192.

——, 1973: Atmospheric ozone and the movement of air in the stratosphere. *Pure Appl. Geophys.*, **106–108**, 1520–1530.

——, D. N. Harrison and J. Lawrence, 1929: Measurements of the amount of ozone in the earth's atmosphere and its relation to other geophysical conditions, part III. *Proc. Roy. Soc., London*, **A122**, 456–486.

Dütsch, H. V., 1969: Atmospheric ozone and ultraviolet radiation. *World Survey of Climatology*, Vol. 4, D. F. Rex, Ed., Elsevier, 383–432.

Harwood, R. S., and J. A. Pyle, 1977: Studies of the ozone budget using a zonal mean circulation model and linearized photochemistry. *Quart. J. Roy. Meteor. Soc.*, **103**, 319–343.

Hunt, B. G., 1969: Experiments with a stratospheric general circulation model. III. Large-scale diffusion of ozone including photochemistry. *Mon. Wea. Rev.*, **97**, 287–306.

——, and S. Manabe, 1968: Experiments with a stratospheric general circulation model II. Large-scale diffusion of tracers in the stratosphere. *Mon. Wea Rev.*, **96**, 503–529.

Johnston, H., and G. Whitten, 1973: Instantaneous photochemical rates in the global stratosphere. *Pure Appl. Geophys.*, **106–108**, 1468–1489.

Katayama, A., 1972: A simplified scheme for computing radiative transfer in the troposphere. Tech. Rep. No. 6, Department of Meteorology, University of California, Los Angeles, 77 pp.

Krueger, A. J., D. F. Heath and C. L. Mateer, 1973: Variations in the stratospheric ozone field inferred from Nimbus satellite observations. *Pure Appl. Geophys.*, **106–108**, 1255–1263.

Lamb, H. H., 1972: *Climate, Present, Past and Future*. Methuen, 613 pp.

London, J., and J. Park, 1973: Application of general circulation models to the study of stratospheric ozone. *Pure Appl. Geophys.*, **106–108**, 1611–1617.

——, and ——, 1974: The interaction of ozone photochemistry and dynamics in the stratosphere. A three-dimensional atmospheric model. *Can. J. Chem*, **52**, 1599–1609.

——, R. D. Bojkov, S. Oltmans and J. I. Kelley, 1976: *Atlas of the Global Distribution of Total Ozone, July 1957–June 1967*. TN113+STR, National Center for Atmospheric Research, Boulder, 276 pp.

Mahlman, J. D., 1972: Preliminary results from a three-dimensional general-circulation model. *Proc. Second Conf. Climatic Impact Assessment Program*, U. S. Department of Transportation Rep. No. DOT-TSC-OST-73-4, 321–337.*

Manabe, S., and J. D. Mahlman, 1976: Simulation of seasonal and interhemispheric variations in the stratospheric circulation. *J. Atmos. Sci.*, **33**, 2187–2217.

McElroy, M. B., S. C. Wofsy, J. E. Penner and J. C. McConnell, 1974: Atmospheric ozone: Possible impact of stratospheric aviation. *J. Atmos. Sci.*, **31**, 287–303.

Mintz, Y., and M. Schlesinger, 1975: Ozone production and transport with the UCLA general circulation model. *Proc. Fourth Conf. Climatic Impact Assessment Program*, U. S. Department of Transportation Rep. No. DOT-TSC-OST-75-38, 201–222.*

Miyakoda, K., L. Umscheid, D. H. Lee, J. Sirutis, R. Lusen and F. Pratte, 1976: The near-real-time, global, four-dimensional analysis experiment during the GATE period. *J. Atmos. Sci.*, **33**, 561–591.

Murcray, D. G., A. Goldman, W. J. Williams, F. H. Murcray, J. N. Brooks, J. Van Allen, R. N. Stocker, J. J. Kosters and D. B. Barker, 1974: Recent results of stratospheric trace-gas measurements from balloon-borne spectrometers. *Proc. Third Conf. Climatic Impact Assessment Program*, U. S. Department of Transportation Rep. No. DOT-TSC-OST-74-15, 184–192.*

* Available from the U. S. Department of Transportation, Federal Aviation Administration, 800 Independence Ave., S.W., AEE-10, Washington, D. C., 20591.

Murgatroyd, R. J., 1969: The structure and dynamics of the stratosphere. *The Global Circulation of the Atmosphere*. G. A. Corby, Ed., Roy. Meteor. Soc., 159–195.

Newell, R. E., 1961: The transport of trace substances in the atmosphere and their implications for the general circulation of the stratosphere. *Geofis. Pura Appl.*, **49**, 137–158.

——, 1963a: Transfer through the tropopause and within the stratosphere. *Quart. J. Roy. Meteor. Soc.*, **89**, 167–204.

——, 1963b: The general circulation of the atmosphere and its effects on the movement of trace substances. *J. Geophys. Res.*, **68**, 3949–3962.

——, 1964a: Stratospheric energetics and mass transport. *Pure Appl. Geophys.*, **58**, 145–156.

——, 1964b: Further ozone transport calculations and the spring maximum in ozone amount. *Pure Appl. Geophys.*, **59**, 191–206.

——, J. M. Wallace and J. R. Mahoney, 1966: The general circulation of the atmosphere and its effects on the movement of trace substances. Part 2. *Tellus*, **18**, 363–380.

——, G. J. Boer and T. G. Dopplick, 1973: Influence of the vertical motion field on ozone concentration in the stratosphere. *Pure Appl. Geophys.*, **106–108**, 1531–1541.

——, J. W. Kidson, D. G. Vincent and G. J. Boer, 1972: *The General Circulation of the Tropical Atmosphere and Interactions with Extratropical Latitudes*, Vol. 1. The MIT Press, 258 pp.

——, J. W. Kidson, D. G. Vincent and G. J. Boer, 1974: *The General Circulation of the Tropical Atmosphere and Interactions with Extratropical Latitudes*, Vol. 2. The MIT Press, 371 pp.

Newson, R. L., 1974: An experiment with a tropospheric and stratospheric three-dimensional general-circulation model. *Proc. Third Conf. Climatic Impact Assessment Program*, U. S. Department of Transportation, Rep. No. DOT-TSC-OST-74-15, 461–473.*

Normand, C., 1953: Atmospheric ozone and the upper-air condition. *Quart. J Roy. Meteor Soc.*, **74**, 39–50.

Ozone Data for the World, 1960–72: Compiled by Canadian Department of Transportation in cooperation with WMO.

Randall, D. A., 1976: The interaction of the planetary boundary layer with large-scale circulations. Ph.D. dissertation, University of California, Los Angeles, 247 pp.

Schlesinger, M. E., 1976a: A numerical simulation of the general circulation of atmospheric ozone. Ph.D. dissertation, University of California, Los Angeles, 376 pp.

——, 1976b: A fast numerical method for explicit integration of the primitive equations near the poles. P-5507, The Rand Corporation, Santa Monica, 50 pp.

Schutz, C., and W. L. Gates, 1971: *Global Climatic Data for Surface, 800 mb, 400 mb: January*. R-915-ARPA, The Rand Corporation, Santa Monica, 173 pp.

Somerville, R. C. J., P. H. Stone, M. Halem, J. E. Hansen, J. S. Hogan, L. M. Druyan, G. Russell, A. A. Lacis, W. J. Quirk and J. Tenenbaum, 1974: The GISS-Model of the global atmosphere. *J. Atmos Sci.*, **31**, 84–117.

Taljaard, J. J., H. van Loon, H. L. Crutcher and R. L. Jenne, 1969: *Climate of the Upper Air: Southern Hemisphere*, Vol. 1. *Temperatures, Dew Points and Heights at Selected Pressure Levels*. NAVAIR Publ. 50-1C-55, Washington, D. C. [Available from U. S. Govt. Printing Office.]

Tiefenau, H., and P. Fabian, 1972: The specific ozone destruction at the ocean surface and its dependence on horizontal wind velocity from profile measurements. *Arch. Meteor. Geophys. Bioklim.*, **A21**, 399–412.

U. S. Standard Atmosphere Supplements, 1966: U. S. Govt. Printing Office, Washington, DC, 289 pp

Washington, W. M., B. Otto-Bliesner and G. Williamson, 1977: January and July simulation experiments with the 2.5° latitude-longitude version of the NCAR general circulation model. NCAR Tech. Note, NCAR/TN-123+STR, 61 pp.

Paper 16

NASA Conference Publication 2076 – Fourth NASA Weather and Climate Program Science Review, January 24-25, 1979, NASA/GSFC, Greenbelt, MD.

Paper No. 32

ADVANCED METEOROLOGICAL TEMPERATURE SOUNDER (AMTS) SIMULATIONS

J. Susskind, *Goddard Space Flight Center, Greenbelt, Maryland,* A. Rosenberg, *Sigma Data Services Corporation, c/o Goddard Space Flight Center, Greenbelt, Maryland* and L. D. Kaplan, *Goddard Space Flight Center, Greenbelt, Maryland*

ABSTRACT

This is a report of simulation studies on temperature retrievals from AMTS and their effect on atmospheric analysis. Observations are simulated from radiosonde reports and observed cloud cover. Temperature retrievals are performed and RMS temperature and thickness errors are calculated relative to the radiosonde profiles and compared to similarly generated HIRS statistics. Significant improvement over HIRS is found throughout the atmosphere but especially in the stratosphere and lower troposphere.

Paper 17

NASA Conference Publication 2076 – Fourth NASA Weather and Climate Program Science Review, January 24-25, 1979, NASA/GSFC, Greenbelt, MD.

Paper No. 61

A GCM SIMULATION OF THE EARTH-ATMOSPHERE RADIATION BALANCE FOR WINTER AND SUMMER

M. L. C. Wu, *National Research Council, Washington, DC and Goddard Space Flight Center, Greenbelt, Maryland*

ABSTRACT

The radiation balance of the earth-atmosphere system simulated by using the general circulation model (GCM) of the Laboratory for Atmospheric Sciences (GLAS) is examined in regards to its geographical distribution, zonally-averaged distribution, and global mean. Most of the main features of the radiation balance at the top of the atmosphere are reasonably simulated, with some differences in the detailed structure of the patterns and intensities for both summer and winter in comparison with values as derived from Nimbus and NOAA (National Oceanic and Atmospheric Administration) satellite observations. Both the capability and defects of the model are discussed.

III. UPPER ATMOSPHERE BRANCH

Paper 18

From the International Union of Geodesy and Geophysics (IUGG) Proceedings – Middle Atmosphere Symposium, December, 1979, Canberra, Australia.

MIDDLE ATMOSPHERE RESPONSE TO MEASURED RELATIVISTIC ELECTRONS

J. R. Barcus (Dept. of Physics, University of Denver, Denver, Colorado 80210 and NAS-NRC Senior Resident Associate, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771 USA)

R. A. Goldberg and E. R. Hilsenrath (both at NASA/Goddard Space Flight Center, Laboratory for Planetary Atmospheres, Greenbelt, Maryland 20771 USA)

J. D. Mitchell (Dept. of Electrical Engineering, University of Texas at El Paso, El Paso, Texas 79968 USA)

It is now apparent from satellite data that relativistic electrons are discharged from the magnetosphere during most auroral substorm events (e.g., Reagan [1], Thorne [2]). These events occur with varying degrees of intensity and spectral hardness, and with major fluctuations in duration and geographic extent. During two independent rocket programs, Aurorozone I and II at Poker Flat Research Range, Alaska (65.1°N, 147.5°W, magnetic dip = 77.2°) in September 1976 and March 1978 respectively, we have observed frequent penetration and absorption of relativistic electrons between 40-70 km altitude. We have also been able to compare this relativistic electron source with bremsstrahlung x-ray energy deposition in the same altitude domain. Correlated atmospheric electrical response was simultaneously observed. Ozone was also measured following each event to determine its departure from quiet nighttime values. The preliminary results for Aurorozone I have already been reported [3]. There we found that the relativistic electrons observed on September 23, 1976 dominated x-ray bremsstrahlung as an energy source to an atmospheric depth of 55 km. Above this height, the deposition ratio rapidly grew to a value greater than 100, clearly demonstrating the dominant role of relativistic electrons as an important energy source for this lower mesospheric domain. This note concentrates on our preliminary findings for Aurorozone II, during which we flew a newly developed payload (XRG, first described in [3]), to make simultaneous measurements of the energy sources and atmospheric electrical response on a single, integrated, parachute-borne payload.

Reprinted from *Applied Optics*, Vol. 18, No. 20, October 15, 1979.

Remote measurement of tropospheric ozone

Jack L. Bufton, Richard W. Stewart, and Chi Weng

Chi Weng is with Systems & Applied Sciences Corporation, Riverdale, Maryland 20840; the other authors are with NASA Goddard Space Flight Center, Greenbelt, Maryland 20771.

Received 16 September 1978.
0003-6935/79/203363-02$00.50/0.

Differential absorption lidar using a pulsed CO_2 laser and a direct detection receiver is capable of significantly improving the existing data base on the tropospheric ozone (O_3) burden. As a ground-based system the lidar could obtain urban to regional scale O_3 measurements with a vertical or horizontal resolution of at least 1 km in the troposphere. As a space-based system the lidar could obtain global scale coverage of the O_3 burden below the stratospheric maximum of O_3. Measurements in both these categories are required to shed light on the dynamics and chemistry of tropospheric ozone.[1]

The CO_2 differential absorption lidar (DIAL) measures backscattered laser energy at two wavelengths in the 9.4-μm band of carbon dioxide, which overlap the 9.6-μm band of ozone. The logarithmic ratio of backscattered laser energy at the two wavelengths is proportional to the integrated absorption by O_3. Knowledge of the absorption coefficient along the propagation path permits an inversion of the energy ratio data to obtain O_3 concentrations. A remote target is used to establish a two-way long path absorption geometry for these measurements.

Atmospheric aerosols serve as the remote target for the ground-based version of this system. The presence of aerosols throughout the troposphere will permit O_3 concentration measurements as desired. This is accomplished by adjusting the starting time for integration of detector output after the laser pulse is transmitted. Only that backscattered laser energy lying within a selected range cell (integration period) contributes to the differential absorption measurement. Since absorption at a given altitude depends on wavelength and the thermodynamic state of the atmosphere, in addition to O_3 concentration, inversion of the lidar data requires models or data for temperature and pressure in addition to an O_3 line parameter compilation.

For space-based applications the ocean surface provides the remote target for the CO_2 DIAL system. The enhanced specular return from the ocean at a nadir viewing angle provides the SNR necessary for energy measurements.[2] Since the main concentration of O_3 is in the stratosphere, a selective weighting must be used to separate the tropospheric component. The location of CO_2 laser lines in the pressure-broadened wings of O_3 lines provides this weighting. Near the earth's surface the O_3 linewidth is about 0.22 cm^{-1}, and quite a number of O_3 lines contribute to absorption at any particular CO_2 line. The linewidth decreases to about 0.05 cm^{-1} and 0.01 cm^{-1}, respectively, at the tropopause (12 km) and the stratospheric maximum of O_3 (22 km). At these higher altitudes relatively few O_3 lines contribute to laser absorption, and those that do have a smaller effect. As a result the absorption coefficient (km^{-1}) is often greater near the earth's surface. Laser lines can be chosen from approximately 15 strong lines in the P branch of the 9.4-μm CO_2 band in order to maximize this weighting to the troposphere. A fine tuning of the weighting effect is possible by control of laser linewidth in the 0.13-cm^{-1} pressure-broadened envelope of the laser. These effects provide a discrete tuning capability to optimize the O_3 measurement even though the CO_2 DIAL system operates at fixed laser frequencies.

The principal components of a prototype CO_2 DIAL system now under development are illustrated in the block diagram of Fig. 1. Output laser energy is 1 J/pulse, multimode, with a 50-nsec pulsewidth and a 1-mrad divergence. This energy is divided more or less equally between two lasing wavelengths selectable by independent grating drives. A small portion of the transmitted laser beam is directed back into the receiver optics for use as an energy monitor for each laser pulse. Receiver optics include a reflector telescope followed by a diffraction grating for separation of the two wavelengths. The detector is a multielement array of HgCdTe photodiodes cooled to 77 K. These elements are operated in the photovoltaic mode, and each has a minimum detectivity of 5×10^{10} cm-$Hz^{1/2}$/W. The detector noise equivalent energy (NEE) is approximately 10^{-16} J for detection of the 50-nsec laser pulses. An electronic data system based on CAMAC and microprocessor technology is used to process detector signals for the differential absorption statistic, $\ln[(E_{R2}/E_{T2})/(E_{R1}/E_{T1})]$, where E_R and E_T are, respectively, the received and transmitted energies measured for laser wavelengths 1 and 2. The data system can record the raw data and invert it to calculate O_3 concentrations. A key feature of the electronics is the capability to perform a gated integration of the received laser energy. This range gating feature, also used in laser ranging systems, acts to reduce the effect of noise and provides a distance measuring and profiling capability for the ground-based measurements.

Backscattered laser energy in both space-based and ground-based systems is calculated at 10^{-14} J/pulse, a factor of 40 dB over the detector NEE. These calculations are based on a target backscatter coefficient of 0.2 sr^{-1} and 10^{-4} sr^{-1},

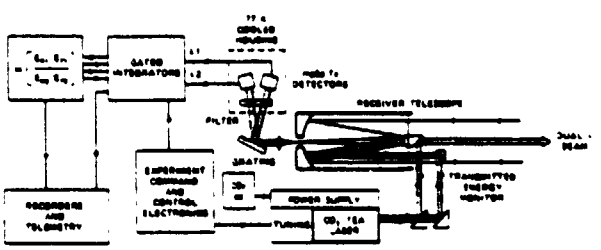


Fig. 1 CO_2 DIAL system components

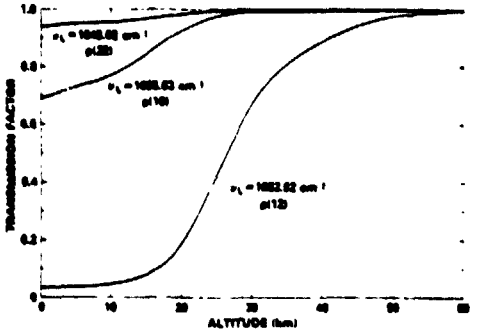


Fig. 2. One-way transmission from space at nadir for three CO_2 laser lines at the indicated center wave numbers ν_L.

respectively, for the ocean surface[2] and a 1-km section of the aerosol laden atmosphere.[3] For the latter case NEE must be increased by a factor of 10 to account for the 7-μsec range gate necessary for an integration over 1 km. The resulting decrease in SNR can be made up by use of a larger diameter collection telescope in the ground-based system. Since a complete differential absorption measurement is made for each dual-wavelength pulse, the SNR can also be improved by a factor of $N^{1/2}$ by averaging over N pulses at the laser firing rate of 2 pulses/sec. Detector SNR is only one of several considerations affecting DIAL measurement accuracy. Kildal and Byer[4] and Remsberg and Gordley[5] have discussed these in detail. It is likely that error in the *a priori* knowledge of O_3 absorption coefficients and the effects of interfering species, especially water vapor, could establish the limiting accuracy in the CO_2 DIAL system described here. It will be necessary to choose laser lines that optimize the differential absorption measurement in the presence of all noise sources.

The space-based measurements result in the O_3 total burden in a vertical column weighted more or less heavily as described to the atmosphere below the O_3 maximum. The choice of laser lines determines the weighting function. This is illustrated in Fig 2, where transmission vs altitude is plotted for three CO_2 laser lines in the 9.4-μm band. The AFGL[6] HITRAN computer code, line parameter compilation, and Mid-Latitude Winter O_3 model were used to compute these one-way transmissions. The $P(22)$ line is weakly absorbed by O_3 and would serve as a good reference wavelength. The $P(12)$ line is heavily absorbed by O_3 with most of the absorption in the stratosphere. This would be an ideal line for total burden measurement except that the two-way transmission would yield a small SNR. The $P(10)$ line would be a good choice for tropospheric O_3 measurement. Approximately $\frac{1}{3}$ of its absorption occurs below 12 km, although only $\frac{1}{10}$ of the O_3 is located in that region. Total transmission on a two-way path is about 50% for this line, which is an adequate signal level. Inversion of the space-based data proceeds by use of vertical profiles of O_3 and interfering species absorption line parameters and the assumption of a shape function for the vertical distribution of O_3. The inversion algorithm can be used to extract a scaling factor for the O_3 total burden in the lower atmosphere. The addition of independent satellite measurements of stratospheric O_3 and meteorological profiles would improve the accuracy of this inversion.

The CO_2 DIAL system as described here differs from earlier systems described by Murray[7] and Asai and Igarashi[8] in that simultaneous energy measurements are conducted at both wavelengths in the differential absorption method. This has the obvious advantage of removing any wavelength-independent noise such as atmospheric turbulence-induced scintillations and target albedo fluctuations. This will greatly aid the ground-based measurements where turbulence is strong and aerosol structure is highly variable. Simultaneous measurements will be an essential feature of a space-based instrument since the nadir point typically moves at about 10 km/sec. No two successive DIAL measurements would use the same target. The range gating feature is also a major improvement that acts to reduce background and detector noise up to 3 orders of magnitude from continuous wave (cw) measurements. Another important technique is the use of the same detector and electronics for measurement of transmitted and received laser energies. Normalization by the transmitted energy acts to remove any pulse-to-pulse variations and any $1/f$ noise slower than the pulse two-way propagation time.

The direct detection scheme considered here would at first appear to be quite inferior to a heterodyne detector. Thermal noise limits give a factor of at least 10^6 advantage to heterodyne systems which can operate at the quantum limit. Menzies[9] has demonstrated CO_2 DIAL measurements with cw lasers and heterodyne detectors. A typical heterodyne efficiency is, however, only a few percent.[10] In addition, spaceborne cw CO_2 lasers are limited to a few tens of watts if they must also have the single-mode operation and stability required in heterodyne detectors. The pulsed CO_2 TEA laser produces about 20 MW in its 50-nsec pulse; hence, the relative advantage of the cw/heterodyne system is lost when the SNR is calculated. Another serious shortcoming of the heterodyne detector is its requirement for a diffraction-limited field-of-view (tens of microradians) and consequent sensitivity to coherent fading (speckle) effects. Gardner[11] has shown that the milliradian field-of-view and transmitter divergence angles possible when using direct detection can produce several orders of magnitude reduction in speckle effects. The combination of a pulsed CO_2 laser and direct detection receiver appears to have superior performance for the space-based measurement of tropospheric ozone.

References

1. R. W. Stewart, S. Hameed, and J. P. Pinto, J. Geophys. Res. **82**, 1134 (1977).
2. K. J. Petri, IEEE Trans. Geosci. Electron. **GE-15**, 87 (1977).
3. R. M. Huffaker, Ed., NOAA Technical Memorandum ERL-WPL-47, Wave Propagation Laboratory, Boulder, Colo. (September 1978).
4. H. Kildal and R. L. Byer, Proc. IEEE **59**, 1644 (1971).
5. E. E. Remsberg and L. L. Gordley, Appl. Opt. **17**, 624 (1978).
6. R. A. McClatchey *et al.*, AFCRL Report 72-0497, Bedford, Mass. (August 1972).
7. E. R. Murray, Opt. Eng. **17**, 294 (1977).
8. K. Asai and T. Igarashi, J. Appl. Phys. Suppl. **14-1**, 137 (1975).
9. R. T. Menzies, Opt. Eng. **17**, 44 (1978).
10. P. Kostuk, M. J. Mumma, M. M. Abbas, and D. Buhl, Infrared Phys. **16**, 61 (1976).
11. C. S. Gardner, RRL Publication 488, U. Illinois, Urbana (Dec. 1977).

Paper 20

VOL. 84, NO. C7 JOURNAL OF GEOPHYSICAL RESEARCH JULY 20, 1979

Comparison of Seasonal Variations of Upper Stratospheric Ozone Concentrations Revealed by Umkehr and Nimbus 4 BUV Observations

JOHN J. DELUISI

NOAA Environmental Research Laboratories, Air Resources Laboratory–GMCC, Boulder, Colorado 80302

CARLTON L. MATEER

Atmospheric Environment Service, Downsview, Ontario, Canada M3H 5T4

DONALD F. HEATH

NASA Goddard Space Flight Center, Greenbelt, Maryland 20771

This paper reports the results of a comparison between upper stratospheric ozone concentration profiles in the region between 22 and 1.4 mbar, as determined from surface-based Umkehr observations and satellite Nimbus 4 BUV observations. The Umkehr data, consisting of monthly averages of observations extending over several years or longer, were obtained at three stations located in the northern hemisphere and two in the southern hemisphere. The BUV data were obtained during the period from May 1970 to March 1971. Aside from some bias in the magnitudes of the Umkehr and BUV data, marked annual cycles of ozone concentration in the upper stratosphere are clearly revealed. Above 4 mbar the profiles show a summer minimum and a winter maximum, while below 4 mbar the annual variation is reversed from this pattern. In the northern hemisphere the winter maximum is accompanied by a secondary minimum of 1- to 2-month duration near 3 mbar. This short-term minimum is much less obvious in the southern hemisphere data. Some of the problem of attempting to monitor long-term changes in the upper stratosphere are discussed briefly.

1. INTRODUCTION

Remote sensing observations of the vertical distribution of ozone using skylight measurements of scattered solar ultraviolet radiation originated in 1934. *Götz et al.* [1934] were the first to infer the main features of an ozone profile from the *Götz* [1931] Umkehr (or reversal) effect. Not long after, Umkehr observations were started at a number of stations where Dobson spectrophotometers were in routine use measuring total columnar ozone. However, only a few of these stations have carried on a program of routine measurements for 10 successive years or longer.

Until the Nimbus 4 backscattering ultraviolet (BUV) experiment began in April 1970 [*Heath et al.*, 1973], Umkehr observations provided the only record, by virtue of their number, of the vertical distribution of ozone, having a potential for determining ozone trends in the upper stratosphere [e.g., *Angell and Korshover*, 1978].

Both Umkehr data [*Rangarajan*, 1969] and BUV data [*Krueger et al.*, 1973] display quite similar seasonal variations. A preliminary comparison of a limited number of Umkehr and BUV observations was done by *DeLuisi and Nimira* [1977]. Both types of observation suggested winter maximum and summer minimum cycles in ozone concentration above 4 mbar, while below this level the cycle appeared to be approximately 6 months out of phase. Also, the relative amplitude of the cycle was seen to diminish with decreasing altitude, i.e., going downward from the highest levels observed ($\gtrsim$ 45–50 km).

Ultraviolet remote sensing methods for observing the vertical distribution of ozone cannot reveal detailed structure owing to certain inherent physical limitations dictated by the natural process of diffuse transmission and reflection (backscattering) by the earth's atmosphere. Basically, ultraviolet light scattered by the atmosphere originates from a rather thick layer with a half-amplitude thickness of approximately 10 km. The altitude of this scattering layer depends mainly upon the solar zenith angle, the strengths of molecular scattering and ozone absorption (which are wavelength dependent), and the vertical profile of ozone [*Mateer*, 1965]. In the case for downward scattering towards the earth's surface an additional scattering layer having its maximum at the surface is a persistent feature which arises from higher-order scattering by the atmosphere.

In this paper we briefly describe the Umkehr and the BUV methods. We then compare monthly averages of Umkehr data with monthly averages of preliminary BUV data. The BUV data were obtained during the period from May 1970 to March 1971, the first year of operation of the Nimbus 4 satellite. Umkehr data are averages of from 8 to 21 years of observations at various stations. The BUV data are preliminary in that all known instrumental corrections have not been applied to the radiance observations, and, moreover, a final measurement evaluation procedure, which will be uniformly applied to all observations, has not been used to generate the present data set.

A problem that frequently prevails with remote sensing methods is that of systematic bias in the estimation (or solution) procedure. The causes can be instrumental or due to the lack of a complete understanding of the physical processes involved. There may also be one or more fundamental limitations in the method itself. The final part of the present investigation looks briefly at the problem of systematic differences between the Umkehr and BUV data.

2. THE UMKEHR METHOD

The Umkehr effect and methods for retrieving the vertical ozone profile from observations of the effect have been described extensively in literature [*Götz*, 1931, *Götz et al.*, 1934,

Paper number 9C0435.
0148-0227/79/009C-0435$01.00

Dütsch, 1959; *Mateer*, 1965]. For our purposes here it will suffice to reiterate that it is an indirect method and, consequently, because of the inherent nonuniqueness of such methods, the derived ozone profiles depend to some extent on the details of the mathematical inversion scheme used to evaluate the observations. In this study, we have used data from *Ozone Data for the World* [1960]. The standard Umkehr evaluation system [*Mateer and Dütsch*, 1964] currently in use at the World Ozone Data Centre (WODC) was used to process the Umkehr measurements.

Comparison of Umkehr-derived ozone profiles with those obtained concurrently by balloon-borne ozonesondes, up to the average balloon ceiling of about 30 km, show that the smoothing inherent in the indirect method cause Umkehr profiles to have too little ozone near the level of maximum ozone density and too much ozone above and below the maximum [*Bojkov*, 1966; *Craig et al.*, 1967; *Dütsch and Ling*, 1969; *Kulkarni and Pittock*, 1970]. Inadequacies in the current Umkehr evaluation system may also be partially responsible for these differences [*DeLuisi and Mateer*, 1971].

In addition to these profile characteristics that are introduced by the particular properties of the solution method, other biases are present because of the neglect of certain physical effects. The neglect of absorption and scattering by haze particles, for example, results in too low ozone concentrations in the higher layers (30–45 km) of the atmosphere and too high concentrations in the lowest layers [*DeLuisi*, 1969; *DeLuisi et al.*, 1975]. Haze effects could conceivably produce some seasonal variations in the profiles. Moreover, the neglect of the temperature dependence of the ozone absorption coefficients not only produces some overall bias in the results but also could produce a fictitious seasonal variation in the profiles [*DeLuisi*, 1971]. Both types of variation are thought to be of the order of a few percent in the higher layers and about 10% in the lowest layers during unperturbed conditions (i.e., low volcanic activity). In the middle layers the temperature dependence exerts little influence on the results.

The WODC solution model divides the atmosphere into nine layers, the first extending from the earth's surface to 250 mbar, while for the second through ninth layers the pressure at the top of the layer is one half that at the bottom. Although a fixed pressure does not imply a fixed height for each layer, it may be helpful to some readers to note that the altitudes of the midpoints of the Umkehr layers are 25.9, 30.5, 35.3, 40.3, and 45.6 km, going from layer 5 to layer 9 in terms of the *U.S. Standard Atmosphere Supplements* [1962]. The amount of ozone above the top of the ninth layer is a fixed 65% of the amount in the ninth layer. Layer ozone content is described in terms of the average ozone partial pressure.

Because many ozone stations take only occasional Umkehr observations, averages of several years of data are necessary to derive reasonably smooth curves depicting the annual variation. An exception to this is Arosa, which is the only station for which sufficient observations were obtained in 1970 and 1971 to permit a more or less time coincident comparison. Monthly mean values of ozone partial pressures in the nine-layer model have been taken directly from the WODC publications for the stations listed in Table 1.

TABLE 1. Percentage Difference Between Annual Average (Excluding April) Umkehr and BUV Data in the Various Layers

Station	Layer 6	Layer 7	Layer 8	Layer 9
Belsk (52°N)	12.7	11.3	1.4	0.5
Arosa (47°N)	3.2	6.6	−3.0	5.5
Tateno (36°N)	14.8	18.9	10.5	7.7
Aspendale (35°S)	10.8	12.4	[illegible]	10.5
Pretoria (25°S)	11.5	−0.4	−10.7	−5.1

Difference is positive when Umkehr average is higher.

3. The BUV Method

The BUV method for estimating the ozone profile was first suggested by *Singer* [1956]. Early satellite experiments using the method have been reported by *Rawcliffe and Elliott* [1966], *Krasnopol'skiy* [1966], *Iozenas* [1968], *Iozenas et al.* [1969a, b], and *Anderson et al.* [1969]. In the Nimbus 4 satellite experiment [*Heath et al.*, 1973] the high-level ozone profile is inferred from measurements of the solar ultraviolet radiation backscattered in the satellite's nadir direction. The physical basis of this experiment and methods used to infer the ozone profile have been outlined by *Krueger et al.* [1973].

Measurements in up to eight wavelength channels between 255.5 and 305.8 nm are used in the inversion. Only those wavelengths which do not significantly penetrate through the main ozone concentration maximum are used and for these wavelengths a single-scattering model provides adequate accuracy. This means that the lower limit of validity of the derived profiles is somewhere between the level of the main maximum and about 30 km (10 mbar). Comparison of inversion results with concurrent rocket ozone soundings [see *Krueger et al.*, 1973] indicates that the BUV results are too low at high levels (about 1 mbar) and also too low near 10 mbar. At high levels the results are (barely) within the probable error bars of the rocket results. (The rocket data have not been used to 'calibrate' the BUV as *London et al.* [1977] did in the case of O_3 data.) Scattering by aerosols is not accounted for in the inversion model, and this means that the BUV results will be too low whenever significant amounts of aerosols are present between 1 and 10 mbar.

The BUV data used in this paper are taken from unpublished monthly summaries of ozone concentrations at fixed pressure levels (1, 2, 4, 10, 15, and 30 mbar) and at every 10° of latitude from 80°N to 80°S, with the exception, of course, of those months and latitudes when and where polar night conditions prevail. For comparison with the Umkehr data the BUV data were linearly interpolated to the midpoint (re log pressure) of the Umkehr layers. It has to be emphasized that the BUV ozone profile data comprise averages around latitude circles.

4. Comparison of Umkehr and BUV Observations

Figure 1 displays curves of monthly averages of BUV (solid lines) and Umkehr (dashed lines) data for Belsk, Poland (52°N), Arosa, Switzerland (47°N), and Tateno, Japan (36°N), for the various Umkehr layers. In layers 8 and 9 there is a marked annual cycle with a winter maximum and summer minimum, and the amplitude of the cycle increases with latitude. There is a distinct secondary minimum imposed on the broad winter maximum. Indeed, in layer 8 this secondary minimum is sufficiently pronounced that one is inclined to identify a double annual cycle with maxima in February and September–October and minima in May–July and December–January. On the average, the amplitude is greater for the BUV results than for the Umkehr results. Considering that the Umkehr data are long-term averages for a single station, whereas the BUV data are around-the-world averages at the station latitude for a single year, the agreement is remarkably

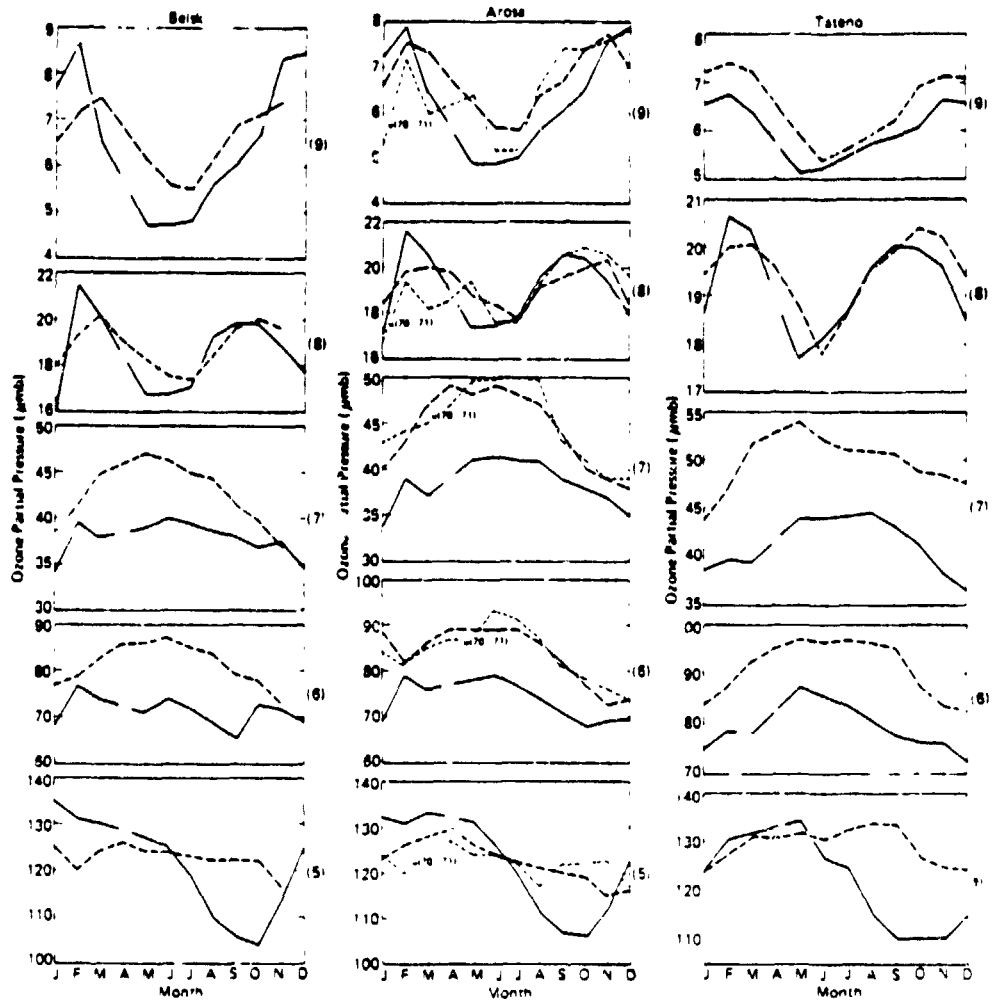


Fig. 1 Plots of monthly average ozone partial pressures over Belsk, Arosa, and Tateno. Standard Umkehr layer numbers are given on the right-hand side. Dashed line is the Umkehr, solid line is the BUV, and lightly dashed line labelled u(70–71) is Umkehr data for the same year as the BUV data. Other Umkehr data are climatological averages for several years or more.

good. The basic winter maximum and summer minimum at these levels are predicted by photochemical theory and arise because of the increased efficiency with temperature of the ozone destruction reaction $O + O_3 \rightarrow 2O_2$ [*Cunnold et al.*, 1975, *Barnett et al.*, 1975; *London et al.*, 1977]. The secondary minimum near 40 km in winter is not explained by pure photochemical theory and it follows that atmospheric transport is probably involved [see *Miller*, 1975]. This phenomenon is at least weakly produced in two-dimensional time-dependent model calculations, which include slantwise eddy transport in the meridional plane. These calculations also produce the transition from the basic winter maximum–summer minimum in the 45- to 50-km region (layer 9) to the summer maximum–winter minimum at lower levels (layers 6–7, see below) as well as the negligible seasonal variation in layer 5. The model has been described by *Yupputun* [1979], the results referenced here are as yet unpublished intermediate results of his model calculations.

The lines of shorter dashes on the Arosa diagrams are Umkehr results for the period April 1970 to March 1971; i.e., they are coincident in time (but not in space) with the BUV data. The shorter-term Umkehr data do not agree as well with the BUV data as do the longer-term Umkehr values. It appears that the longer-term time averaging at the single station is nearly equivalent to space averaging around the latitude circle at these levels in the atmosphere.

It is worth noting that the BUV wavelengths that 'sense' the ozone in these layers have a smaller temperature effect on their absorption coefficients than do the Umkehr wavelengths and that they are essentially unaffected by multiple scattering or by haze in the troposphere and lower stratosphere. Even seasonal variations in the stratospheric Junge aerosol layer should have little impact on the BUV results for layers 8 and 9.

In Umkehr layers 6 and 7 the agreement is not as good. Although both methods suggest roughly an annual cycle with a summer maximum and a winter minimum, there is an obvious bias, with the BUV giving consistently lower values than the Umkehr. In layer 5 the annual cycles are not in agreement. In general, this layer is below, at least in part, the lower validity limit for BUV profiles, and the BUV results tend to follow the first-guess inversion profile, which in this layer is strongly correlated with total ozone. In both Umkehr and balloon soundings, layer 5 ozone content displays no significant annual variation [*Dütsch*, 1966, *Dutsch and Ling*, 1973]

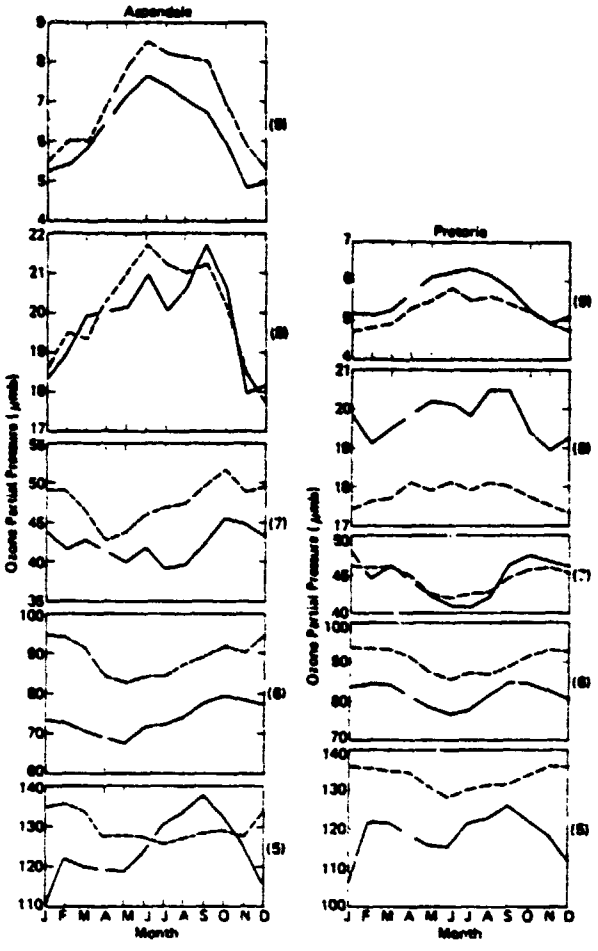


Fig. 2. Same as Figure 1 except fo outhern hemisphere stations Aspendale and Pretoria.

In Figure 2 we present similar comparison curves for Aspendale, Australia (35°S), and Pretoria, South Africa (25°S). In layers 8 and 9, Aspendale follows the broad pattern of Figure 1, with a winter maximum and a summer minimum, but there is only a suggestion of the secondary winter minimum in layer 8 (compare with Tateno at very nearly the same latitude in the northern hemisphere). At Pretoria, closer to the equator, the amplitude of the annual cycle is substantially reduced, and there is only a very weak indication that the secondary winter minimum may be present in layer 8. The much-reduced amplitude of the secondary winter minimum in the southern hemisphere means that the transport processes producing the minimum are probably much weaker in the southern hemisphere than in the northern hemisphere winter season.

There is a relatively large bias between Umkehr and BUV results in layer 8 at Pretoria and, indeed, this is the only case where BUV results are significantly higher than the Umkehr results (some low-latitude northern hemisphere stations also exhibit this bias, but to a lesser extent). The bias disappears in layer 7, and in layer 6, Pretoria exhibits the 'normal' layer 6 bias with Umkehr results higher. It is possible that the 25°S BUV results are in error because of the south Atlantic anomaly in the trapped particle radiation belt. The Nimbus 4 satellite passes through this anomaly, where the energetic particles cause the sapphire window of the BUV photomultiplier to fluoresce, and a spurious component is added to the measurements. The evaluation program discards instrument data when the effect is obvious, but there may be some instrument scans on the fringe of the belt where the effect is large enough to impact on the latitudinal average yet not sufficiently large to cause the evaluating program to discard them. The effect of the error is to produce results that are too high in layer 8.

At both Aspendale and Pretoria the layer 6 and 7 results follow roughly the same annual pattern as observed at the northern hemisphere stations, although there is a tendency for the highest values to appear in the spring. Our main concern here is, however, with the similarity between the annual cycles displayed by the two types of data, and these are roughly similar. In layer 5 the similarities in the annual cycles are no longer apparent, for the reasons given in the discussion of Figure 1.

In Table 1 we compare annual average (excluding April) percentage difference (Umkehr-BUV) for each station for which results are given in Figures 1 and 2. In all layers the Umkehr result is almost invariably higher, the exceptions being layer 8 at Arosa and layers 7, 8, and 9 at Pretoria, where the BUV errors mentioned earlier may be responsible. It may be significant that Arosa, where the overall differences are least, is the only station at which haze/cloud corrections are applied to the Umkehr observations before the ozone profile is estimated. In view of the possible bias errors inherent in the Umkehr results discussed earlier in Section 2 and the fact that all known errors in the BUV observations have not been corrected for, the existence of the bias is not surprising.

5. Concluding Remarks

Upper stratospheric ozone profiles derived from both BUV and Umkehr data display similar marked seasonal variations of about the same phase and magnitude in the 38- to 50-km region. Between 28 and 38 km the seasonal variations are less marked, but the same rough picture emerges for both data sets.

On the basis of this evidence, if both data types indicate a decreasing (or increasing) trend over a period of years, may we conclude that such a trend indeed exists? We are tempted to argue in the affirmative. However, unless there is a separate means for monitoring dust, especially stratospheric dust, one cannot be certain that an apparent change in ozone is not really just a change in the dust content, with ozone remaining constant. Certainly, one need only examine the Umkehr results for Aspendale for layer 9 for 1962–1963 [*Ozone Data for the World*, 1960] to convince himself that Umkehr data alone cannot be depended upon for such inferences.

On the other hand, since the stratosphere well above the Junge layer seems less likely to be affected by volcanic debris, BUV data should be superior to Umkehr data for monitoring trends in the 38- to 50-km range, provided that the calibration problems of flying such a monitoring instrument in space can be overcome. The final processing of Nimbus 4 BUV profile data, which will shortly be get. g underway, may shed some ght on these matters.

Acknowledgments. We are indebted to J. Angell for reviewing the manuscript and providing helpful comments. We are grateful to R. K. R. Vupputuri for kindly making available intermediate results of his model calculations.

References

Anderson, G. P., C. A. Barth, F. Cayla, and J. London, Satellite observations of the vertical ozone distribution in the upper atmo-

sphere, Proc. Symp. on Atmos. Ozone, Monaco, 1968, *Ann. Geophys.*, *25*, 341, 1969.

Angell, J. K., and J. Korshover, Global ozone variations: An update into 1976, *Mon. Weather Rev.*, *106*, 725–737, 1978.

Barnett, J. J., J. T. Houghton, and J. A. Pyle, The temperature dependence of the ozone concentration near the stratopause, *Quart. J Roy Meteorol. Soc.*, *101*, 245–257, 1975.

Bojkov, R. D., Differences in the vertical ozone distribution deduced from Umkehr and ozonesonde data at Goose Bay, *J. Appl. Meteorol.*, *5*, 872–877, 1966.

Craig, R. A., J. J. DeLuisi, and I. Stuetzer, Comparison of chemiluminescent and Umkehr observations of ozone, *J. Geophys. Res.*, *72*, 1661–1665, 1967.

Cunnold, D., F. Alyea, N. Phillips, and R. Prinn, A three-dimensional dynamical-chemical model of atmospheric ozone, *J. Atmos. Sci.*, *32*, 170–194, 1975.

DeLuisi, J. J., A study of the effect of haze upon Umkehr measurements, *Quart. J. Roy. Meteorol. Soc.*, *95*, 181–186, 1969.

DeLuisi, J. J., Effect of the temperature dependence of ozone absorption on vertical ozone distribution deduced from Umkehr observations, *J. Geophys. Res.*, *76*, 2131–2137, 1971.

DeLuisi, J. J., and C. L. Mateer, On the application of the optimum statistical inversion technique to the evaluation of Umkehr observations, *J. Appl. Meteorol.*, *10*, 328–334, 1971.

DeLuisi, J. J., and J. Nimira, Preliminary comparison of satellite BUV and surface-based Umkehr observations of the vertical distribution of ozone in the upper stratosphere, *J. Geophys. Res.*, *83*, 379–384, 1977.

DeLuisi, J. J., B. M. Herman, R. S. Browning, and R. K. Sato, Theoretically determined multiple-scattering effects of dust on Umkehr observations, *Quart. J. Roy. Meteorol. Soc.*, *101*, 325–331, 1975.

Dütsch, H. U., Vertical ozone distribution over Arosa, final report, Ozone and General Circulation in the Stratosphere, *Rep. AFCRL-TR-59-290*, 39 pp., Air Force Cambridge Res. Lab., Bedford, Mass., 1959.

Dütsch, H. U., Two years of regular ozone soundings over Boulder, Colorado, *Tech. Note NCAR-TN-10*, 443 pp., Nat. Center for Atmos. Res., Boulder, Colo., 1966.

Dütsch, H. U., and C. Ling, Determination of vertical ozone distribution, *Ann. Geophys.*, *25*(1), 211–214, 1969.

Dütsch, H. U., and C. Ling, Fourteen-year series of vertical ozone distribution over Arosa, Switzerland, from Umkehr measurements, *Pure Appl. Geophys.*, *106–108*, 1139–1150, 1973.

Götz, F. W. P., Zum Strahlungsklima des Spitzbergen Sommers, *Gerlands Beitr. Geophys.*, *31*, 119–154, 1931.

Götz, F. W. P., A. R. Meetham, and G. M. B. Dobson, The vertical distribution of ozone in the atmosphere, *Proc. Roy. Soc. London.*, *A145*, 416–446, 1934.

Heath, D. F., C. L. Mateer, and A. J. Krueger, The Nimbus-4 backscatter ultraviolet (BUV) atmospheric ozone experiment, Two years' operation, *Pure Appl. Geophys.*, *106–108*, 1238–1253, 1973.

Iozenas, V. A., Determining the vertical ozone distribution in the upper atmospheric layers from satellite measurements of ultraviolet solar radiation scattered by the earth's atmosphere, *Geomagn. Aeron.*, *8*, 403–407, 1968.

Iozenas, V. A., V. A. Krasnopol'skiy, A. P. Kuznetsov, and A. I. Lebedinskiy, Studies of the earth's ozonosphere from satellites, *Izv. Atoms. Oceanic Phys.*, *5*, 77–82, 1969*a*.

Iozenas, V. A., V. A. Krasnopol'skiy, A. P. Kuznetsov, and A. I. Lebedinskiy, An investigation of the planetary ozone distribution from satellite measurements of ultraviolet spectra, *Izv. Atmos. Oceanic Phys.*, *5*, 219–223, 1969*b*.

Krasnopol'skiy, V. A., The ultraviolet spectrum of solar radiation reflected by the terrestrial atmosphere, and its use in determining the total content and vertical distribution of atmospheric ozone, *Geomagn. Aeron.*, *6*, 236–242, 1966.

Krueger, A. J., D. F. Heath, and C. L. Mateer, Variations in the stratospheric ozone field inferred from Nimbus satellite observations, *Pure Appl. Geophys.*, *106–108*, 1254–1263, 1973.

Kulkarni, R. N., and A. B. Pittock, Results of a comparison between Umkehr and ozonesonde data, *Quart. J. Roy. Meteorol. Soc.* *96*, 739–743, 1970.

London, J., J. E. Frederick, and G. P. Anderson, Satellite observations of the global distribution of stratospheric ozone, *J. Geophys. Res.*, *82*, 2543–2556, 1977.

Mateer, C. L., On the information content of Umkehr observations, *J. Atmos. Sci.*, *22*, 370–381, 1965.

Mateer, C. L., and H. U. Dütsch, Uniform evaluation of Umkehr observations from the world ozone network, in *Proposed Standard Umkehr Evaluation Technique*, 105 pp, National Center for Atmosphere Research, Boulder, Colo., 1964.

Miller, A. J., Meteorological studies based on satellite ozone data, Satellite workshop 2, *Rep. NCAR. CQ-4 and 1975-ASP*, Nat. Center for Atmos. Res., Boulder, Colo., 1975.

Ozone Data for the World, report, Atmos. Environ. Serv., Toronto, Canada, 1960.

Rangarajan, S., A worldwide anomaly in the concentration of ozone above 40 km, *J. Atmos. Sci.*, *26*, 613–616, 1969.

Rawcliffe, R. D., and D. D. Elliott, Latitude distribution of ozone at high altitudes, Deduced from a satellite measurement of the earth's radiance at 2840 Å, *J. Geophys. Res.*, *71*, 5077–5089, 1966.

Singer, S. F., Geophysical research with artifical earth satellites, *Advan. Geophys.*, *3*, 301–367, 1956.

U.S. Standard Atmosphere Supplements, report, U.S. Govt. Print. Office, Washington, D.C., 1966.

Vupputuri, R. K. R., The structure of the natural stratosphere and the impact of chlorofluoromethanes on the ozone layer investigated in a 2-D time dependent model, *Pure Appl. Geophys.*, *117*, 448–485, 1979.

(Received May 1, 1979,
revised October 13, 1978,
accepted March 5, 1979.)

Paper 21

Tellus (1978), 30, 341–349

Vertical temperature and density patterns in the Arctic mesosphere analyzed as gravity waves

By I. J. EBERSTEIN and J. S. THEON, *Laboratory for Atmospheric Sciences, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, U.S.A.*

(Manuscript received September 26, 1975; in final form January 4, 1978)

ABSTRACT

Three series of rocket soundings were conducted from high latitude sites during winter. In the first series, four pitot pressure soundings were launched during a 13-hour period from Ft. Churchill, Canada (59° N) on January 31, and February 1, 1967. The second series consisted of one pitot sounding and two grenade soundings carried out during a 3-hour period on January 13–14, 1970. Temperature and wind profiles and one density profile were observed independently to obtain the thermodynamic structure, the wind structure, and thus their interdependence in the mesosphere. The third series of soundings was conducted from Point Barrow, Alaska (71° N) on December 6, 1971. This series consisted of five soundings of which the first two and the last two were pitot–grenade pairs. Temperature profiles from all soundings in each series were averaged, and a smooth curve (or series of smooth curves) drawn through the points. A hydrostatic atmosphere based on the average, measured temperature profile was computed, and deviations from the mean atmosphere were analyzed in terms of gravity wave theory. The vertical wavelengths of the deviations were 10–20 km, and the wave amplitudes slowly increased with height. The experimental data were matched by calculated gravity waves having a period ranging between 15 and 80 minutes and horizontal wavelengths of 60 to 280 km. Our interpretation is generally consistent with the results of others who have studied gravity-acoustic waves in the atmosphere. The wind measurements are consistent with the thermodynamic measurements. The results also suggest that gravity waves travelled from East to West with a horizontal phase velocity of approximately 60 m/s.

1. Introduction

The present paper treats several series of rocket soundings of atmospheric temperature, density, and wind structure. The soundings are of two types, pitot and grenade. For the former, a pitot stagnation probe is mounted on a rocket vehicle, whose position and velocity are measured by radar tracking. For an incompressible fluid, we have the simple formula

$$\rho = \frac{2(P^0 - P)}{u^2} \qquad (1)$$

Density is determined if stagnation pressure, static pressure, and velocity are measured. Addition of compressibility makes the analysis somewhat more involved without affecting the basic principle presented above (Horvath et al., 1962). For the grenade experiment, a series of grenades are exploded in the atmosphere, and the position and time of each explosion is determined. The time, and direction of arrival of the spherical sound pulse are measured by a ground based microphone array. This information is used to deduce the mean temperature and horizontal wind speed in atmospheric slabs whose thickness varies between 2 km and 4 km depending on the particular grenade payload used (Nordberg and Smith, 1964).

Essentially, the pitot technique provides a high vertical resolution (0.5 km) direct measurement of atmospheric density, while the grenade technique gives a direct measure of atmospheric temperature and horizontal wind with a 2–4-km vertical resolution. For both techniques, the time to make the atmospheric traverse is approximately 1 minute, resulting in a virtually instantaneous picture of the vertical structure.

2. Outline of gravity wave theory

Waves in a stratified fluid under the influence of gravity appear to have been initially discussed by Burnside (1889) and Love (1891). Both authors treated an incompressible fluid. Görtler (1943) used schlieren photography to show experimentally that disturbances in an incompressible stratified medium under the influence of gravity propagate along characteristic rays. Lamb (1909) treated a compressible, adiabatic, ideal gas whose density is stratified by gravity. Since the early contributions by Burnside, Love, and Lamb, there have been many theoretical papers on various types of gravity waves. Reasonably up-to-date treatments of the subject may be found in Eckart (1960), Yih (1965), and Tolstoy (1973).

The simple theory involves linearizing the atmospheric equations of motion and specifying a sinusoidal solution, i.e.,

$$\phi = \phi_0 \exp\left[-i(\omega t + k_x x + k_z z)\right] \tag{2}$$

Perturbations in pressure, density, temperature, and gravity wave generated wind are related to ϕ by what Hines (1960) has called polarization relations, and a dispersion relation which has the form:

$$n_z^2 = \left\{1 - \left(\frac{\omega_A}{\omega_B}\right)^2\right\} - n_x^2\left\{1 - \left(\frac{\omega_B}{\omega}\right)^2\right\} \tag{3}$$

where

$$n_x = \frac{k_x a}{\omega} \quad n_z = \frac{k_z a}{\omega} \tag{4}$$

(Pitteway and Hines, 1963).

Since atmospheric gradients are usually weak, the ray approximation for vertical wave propagation was used. Volland (1969a) has shown that the ray treatment is a sufficient approximation for gravity waves propagating obliquely upwards, provided the horizontal wave number is in the order of, or greater than ω/a, where a is the sound speed, and ω is the circular frequency of the wave. This statement is equivalent to the requirement that the horizontal phase velocity must be less than the sound speed. For vertical wavelengths up to 20 km, the ratio of phase velocity to sound speed tends to be less than 0.25.

The next problem which comes to mind is how to take account of background wind. The frequency of a moving fluid particle, or the "intrinsic frequency" may be defined as

$$\Omega = \omega - \mathbf{k} \cdot \mathbf{U} \tag{5}$$

where U is wind velocity.

The conditions under which the above procedure is valid have been discussed by Jones (1969), and are essentially satisfied. Experimentally, it turns out to be very difficult to obtain a good value of the background wind U. On one hand, a large number of wind soundings is needed to average out the gravity wave contribution. On the other hand, the velocity U is relative to the layer in which the gravity waves are generated, a fact which adds further complication to Equation (5). Finally, the vector nature of the equation requires some prior knowledge of the direction in which the gravity wave is propagating. However, a phenomenon which was originally viewed as an imperfection in the measurement technique, has turned out to be useful in this respect. The rocket trajectory is at an angle to the vertical, resulting in a horizontal vehicle velocity of some 250 m/s, and a horizontal travel distance of 12 km during the measurement. Examination of the experimental curves shows that the wavelength decreases with altitude. Thus the rocket is going into the wave train. Since the launch direction from Pt. Barrow is due east, we may conclude that the wave train is coming from an easterly direction. The same situation prevails at Ft. Churchill. Our results suggest that the waves in these instances travelled from East to West. Thus it becomes possible to use an est[imate] of wind U, relative to the layer at which the rocket measurement begins, i.e., approximately 35 km.

3. Experimental procedure

The experimental data consist of three winter series of rocket soundings from high latitude sites. In the first series, four pitot soundings were conducted during a 13-hour period from Ft. Churchill, Canada (59° N) spanning January 31 and February 1, 1967 (see Table 1) (Smith et al., 1969). The second series consisted of three soundings carried out during a 3-hour period on January 13–14, 1970 from Churchill (Smith et al., 1972). The first sounding was made with the acoustic

Table 1. *Launch sequence of first series. Fort Churchill (59° N); 1967*

GMT	Date	Time after first shot	Type
2317	Jan 31	0h00m	Pitot
0538	Feb 1	6h21m	Pitot
0826	Feb 1	9h09m	Pitot
1158	Feb 1	12h41m	Pitot

Table 2. *Launch sequence of second series. Fort Churchill (59° N); 1970*

GMT	Date	Time after first shot	Type
2223	Jan 13	0h00m	Grenade
2351	Jan 13	1h28m	Pitot
0135	Jan 14	3h12m	Grenade

Table 3. *Launch sequence of third series. Point Barrow (71° N); 1971*

GMT	Date	Time after first shot	Type
0300	Dec 6	0h00m	Pitot
0302	Dec 6	0h02m	Grenade
0442	Dec 6	1h42m	Pitot
0752	Dec 6	4h52m	Pitot
0802	Dec 6	5h02m	Grenade

grenade technique to measure the temperature and wind profiles in the 35–90 km region; the second sounding was a pitot probe launched 88 minutes later to measure the density profile; and the last sounding employed the acoustic grenade technique 104 minutes after the pitot sounding (see Table 2). The third series of soundings was conducted from Point Barrow, Alaska (71° N) on December 6, 1971 (Smith et al., 1974). This series consisted of five soundings, of which the first two and last two were pitot–grenade pairs. The remaining sounding was performed with a pitot probe. The objective of launching such pairs was to obtain high vertical resolution thermodynamic structure (pitot) and wind information (grenade) simultaneously. The series commenced at 0300 GMT with a pitot probe, which was followed by a grenade at 0302 GMT. These profiles were flown over essentially the same trajectory only 2 minutes apart. The remaining soundings, a pitot at 0442 GMT, a pitot at 0752 Gmt, and a grenade at 0802 GMT completed the series (see Table 3). In the second and third series of soundings, the temperature and wind profiles from the grenade technique and the density profiles from the pitot technique provided independent examination of the thermodynamic structure, the wind structure and the relationship of each in the mesosphere.

4. Analysis of data

Temperature profiles from all soundings in each series were averaged and a smooth curve drawn through the points. A hydrostatically determined atmosphere based on the average measured temperature profile was computed. Deviations from this mean hydrostatic atmosphere were then considered.

The hydrostatic equation and the equation of state are used to obtain the remaining two thermodynamic parameters in both the grenade and pitot techniques. Such an approach assumes that vertical accelerations in the atmospheric motions are negligible relative to the acceleration due to gravity, and is valid for mean atmosphere calculations. Smith et al. (1968) compared the results of grenade soundings and pitot probe soundings at Wallops Island, Va. They found temperature agreement to be better than 3 K below 60 km and better than 5 K between 60 and 90 km altitude, confirming the validity of the hydrostatic approximation for background atmosphere calculations.

For the first test series, the ratio of the density measured by the pitot probe to the hydrostatically computed density (based on the average temperature profile) was plotted as a function of altitude for each probe. The density variations are shown in Fig. 1. Fig. 2 shows the corresponding power spectral density versus wavelength for atmospheric density variations between 28 km and 80 km. Analysis techniques employed are discussed in Bendat and Piersol (1966) and Blackman and Tukey (1958). (In this case "power" is defined in terms of wavelength and amplitude.) Wave-

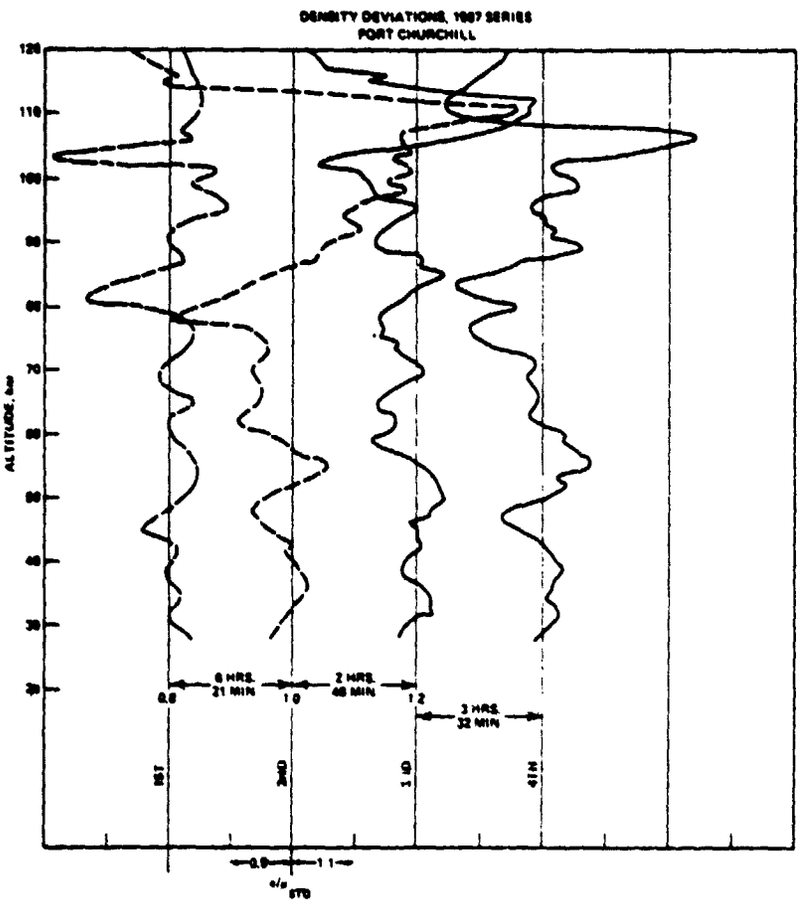


Fig. 1. Measured density deviations for 1967 Ft. Churchill series.

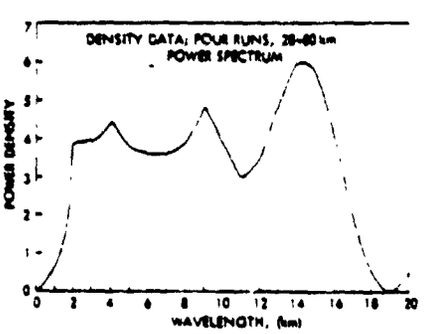


Fig. 2. Power spectral density versus wavelength for experimental deviations, 1967 series.

lengths between 2 km and 17 km are represented, with the primary peak occurring around 14 km, and the secondary peak around 9 km.

For the second test series, which was launched from Churchill (59° N), the observed density perturbation is shown as a solid curve in Fig. 3. The broken curve represents a gravity wave having horizontal wavelength of 60 km and a period of 20 minutes. Fig. 4 compares the wind pattern generated by the above gravity wave with the measurement of the first grenade sounding. Note that the vertical winds include values of some 10 m/s in the 80 km region. Justus and Edwards (1971) have reported measured vertical velocities

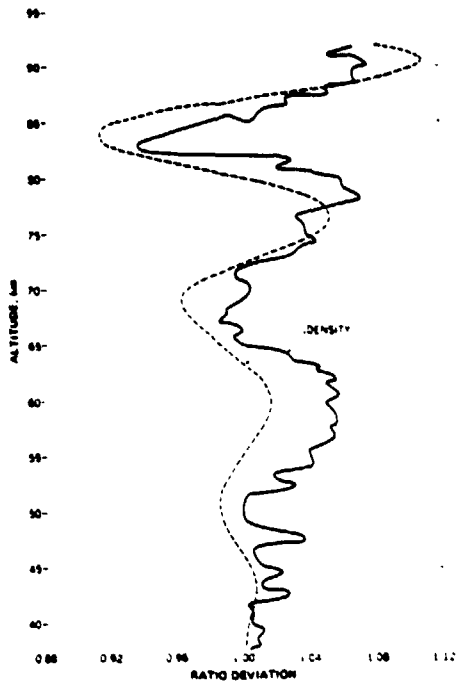


Fig. 3. The observed density perturbations (solid curves) over Churchill on January 13–14, 1970 compared with the perturbations calculated from theory (broken curves) as a function of altitude.

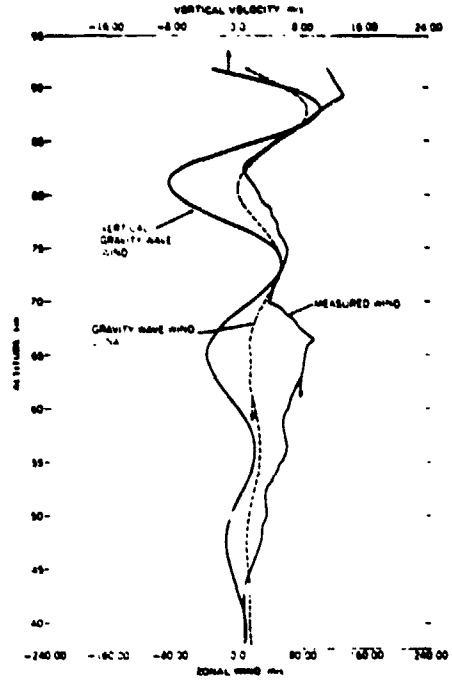


Fig. 4. The observed zonal wind component compared with the gravity wave wind.

on the order of 20 m/s at altitudes between 88–118 km. Although these measurements were made at a different latitude and time of year, they are at least comparable in magnitude. The Justus and Edwards data do confirm the existence of relatively large vertical velocities at high altitudes.

The third set of measurements which consisted of pitot and grenade soundings from Point Barrow (71° N) began with a pitot and a grenade launched 2 minutes apart. The temperature perturbations obtained from the pitot profiles are shown in Fig. 5 together with a theoretical gravity wave which was matched to the first perturbation for a best fit (by eye). The gravity wave structure is based upon the original match of the first pitot and allowed to propagate with time to correspond to the times of the observational data. The best match was found to have a horizontal wavelength of 70 km and a period of about 18 minutes, or 280 km and a period of 72 minutes. It is significant that only the first theoretical curve was fitted to the observed data, and that the remaining matches followed from calculation. Thus our data indicate that not only are the spatial density and wind structure measurements compatible with gravity wave theory, but also the expected behavior with time.

Fig. 6 shows a comparison between the observed wind pattern and the calculated wind pattern associated with the gravity wave matched to the first temperature perturbation. A mean drift wind of 40 m/s has been included. The measured wind is obtained from the grenade sounding, so it represents a series of values in which the winds were averaged over layers (2–3 km thick) between grenade explosions.

5. Discussion

Three series of soundings were carried out at high latitude sites during winter. The first series gave four essentially instantaneous vertical density traverses during a 13-hour period at Ft. Churchill. The density variations from the mean strato-

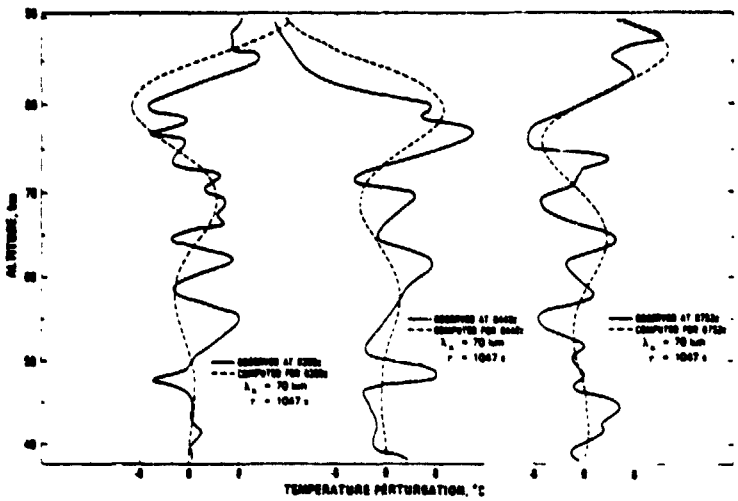


Fig. 5. 1971 series gravity waves matched to pitot temperature deviation.

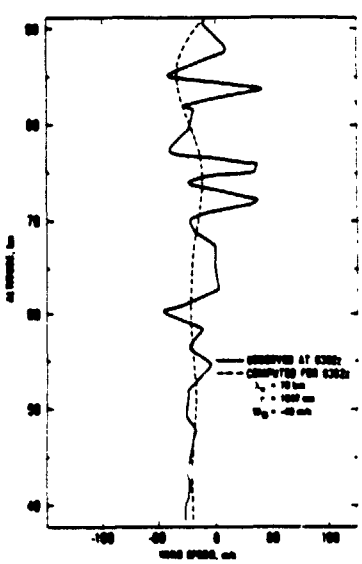


Fig. 6. Comparison between theoretical and observed wind patterns for 1971 series.

spheric conditions were matched by gravity waves having vertical wavelengths of 14–17 km. A reasonable fit to the data was obtained with a plane gravity wave characterized by a horizontal wavelength of 70 km and a period of 18 minutes or 280 km and 70 minutes. The wave amplitude increased with altitude, but at a rate much slower than the exponential growth predicted for undamped waves. Thus one must conclude that the waves were heavily damped as they propagated upward through the atmosphere. The term "damped" in this case is taken to include general wave dissipation by various possible mechanisms, including non-linear effects of the gravity waves themselves. Following Lindzen (1968) we have looked at the Richardson number, and found it to be small or negative between approximately 53 km and 85 km. Thus, it is quite possible that dissipation in this height range is largely the result of non-linear effects in the gravity waves. Such an interpretation is certainly consistent with the observed slow vertical growth rate below 85 km.

Gossard (1962) observed gravity waves in the troposphere, and found the period to vary between 15 minutes and 120 minutes. The horizontal wavelength for waves of 15-minute period was 19 km, and for 2-hour period, it was 150 km. The difference between our observations and those of Gossard may be explained by assuming that tropospheric waves with short vertical wavelengths are damped out. In fact Gossard (1962) found that the maximum leakage into the upper atmosphere occurs at periods of 10 to 20 minutes, starting with a white energy spectrum in the troposphere. Our value of 18 minutes is thus consistent with

Table 4. *Horizontal phase velocity 200° K; 18-minute period*

λ_t (km)	λ_z (km)	V_{pz} (m/s)	V_{pz} (m/s)
5	18.2	16.9	4.6
10	36.4	33.7	9.3
15	54.4	50.4	13.9
20	72.3	67.0	18.5
25	90.0	83.3	23.1

Gossard's prediction. Gossard found a horizontal phase speed of approximately 21 m/s. For an 18-minute period and 200°K, the Table 4 relates our estimated horizontal phase velocity, V_{pz}, to vertical wavelength, λ_z. 200 degrees K was chosen because it is approximately the temperature at the bottom of our test region. Since V_{pz} is essentially λ_z/τ, it remains constant as long as the above ratio remains constant, and is thus equally valid for the 280 km waves.

From noctilucent cloud studies, Witt (1962) observed characteristic wavelengths of 30–40 km, and of 75 km, the latter being very nearly the 70 km deduced in the present study. Witt (1962) also estimated the wave velocity with respect to a frame of reference moving with the cloud system, obtaining a range of 70 to 135 m/s. The agreement with the values given in our table is reasonable. Noctilucent clouds are formed in the 75 to 85 km region of the atmosphere.

Our results are also generally consistent with the discussion of internal atmospheric gravity waves at ionospheric heights presented by Hines (1960). Thus the 12-km vertical wavelength which Hines deduced from the meteor wind data of Greenhow and Neufeld (1955, 1959) is in reasonable agreement with our observations. Hines considered 125 m/s to be a typical horizontal phase speed and quoted the following experimental results. Munro (1958) found phase speeds to generally be in the range of 51–175 m/s, in agreement with both our deductions and Witt's (1962) noctilucent cloud observations. Heisler (1958) found phase speeds to range between 97–207 m/s. The phase speed agreement is remarkable considering that ionospheric heights range up to more than 100 km above the top of our observation regime. Even our longer period of 70 minutes is significantly shorter than the 200 minutes deduced by Hines (1960) from correlation studies of meteor wind trails. However, a careful reading of Hines' paper suggests that the 200 minutes is an upper limit to the period, rather than a most characteristic value.

The results of the third series confirmed the conclusions about vertical wavelength and amplitude of the first two series and the agreement of measured wind with gravity wave predictions. It was also possible to trace one gravity wave pattern through three consecutive samples, indicating that our period is reasonable. However, the fact that the data can be matched by both 70-km and 280-km horizontal wavelengths (corresponding to 18-minute and 70-minute periods, respectively) suggests one of three possibilities, namely (1) the 70-km value represents an aliasing error, which seems unlikely in view of Witt's noctilucent cloud observations, (2) the 70-km horizontal waves represent an interference pattern between waves of 280 km or longer wavelengths, an interpretation consistent with Witt's noctilucent cloud measurements, or (3) the horizontal wave packet contains both 70-km, 18-minute and 280-km, 70-minute waves. The last interpretation disagrees with Hines' 200-minute suggested period but agrees with Gossard's tropospheric measurements. Our data cannot distinguish between explanations "2" and "3". However, Midgley and Liemohn (1966) suggest a wavelength of a few hundred kilometers, essentially in agreement with our 280-km wave. It should be mentioned that an excellent experimental technique exists for studying gravity waves at cloud level, namely observation of sequential satellite photographs. Such pictures should yield both the horizontal wavelength and the horizontal phase velocity relative to the ground.

It might be noted that Lindzen (1971) showed that knowledge of vertical wavelength and period can be used to estimate vertical eddy transport coefficients. Our estimated vertical velocities in the 80-km region are in the order of 2.5 to 10 m/s, compared with 1 m/s assumed by Lindzen from tidal theory. However, he divides his vertical velocity by a number N, characteristic of the particular transport process considered. N is the lifetime of the oscillation in cycles to the nearest integer. From the variability of tides revealed in radio meteor data, Lindzen suggests a value of 10–20 for N. For his 6-hour gravity wave, Lindzen (1971) takes N to be 5. For a 1-hour period, our vertical velocity is 2.5 m/s. Keeping N at 5, and

using Lindzen's formula for estimating eddy diffusivity one obtains

$$D \approx \left(\frac{0.5}{5}\right) \quad (2.5 \times 100\ \text{cm/s})\,(3 \times 10^5\ \text{cm})$$

$$= 7.5 \times 10^6\ \text{cm}^2/\text{s}$$

where

$$D \approx \left(\frac{0.5}{N}\right) \quad (w) \quad \left(\frac{\lambda_z}{2\pi}\right)$$

For the 18-minute waves, the value for D becomes $3.0 \times 10^7\ \text{cm}^2/\text{s}$.

Hodges (1969) has estimated the effective eddy diffusivity associated with gravity waves which reach an amplitude where they generate breakers and thus considerable turbulence. Considering a gravity wave having a vertical wavelength of 16 km, the D_{max} value for $\lambda_z = 70$ km is $7.0 \times 10^7\ \text{cm}^2/\text{s}$, and for $\lambda_z = 280$ km it is $2.0 \times 10^7\ \text{cm}^2/\text{s}$. These values are approximately 2.5 times as large as what one estimates using the Lindzen approximation. However, the agreement must be considered good, considering the uncertainty in the quantity N. Furthermore, the observed waves do have some growth with altitude, while Hodges' value of eddy diffusivity would result in no growth with altitude.

The final mystery is why the waves appear to be coming out of the east, when everybody knows that the prevailing winds are westerly. The answer in this case is quite simply that the prevailing winds above 45 km, i.e., in most of the region considered, are in fact easterly, even though the tropospheric and lower stratospheric winds are westerly. In fact it is quite probable that the gravity waves are generated in a stratospheric shear zone where the winds change from westerly to easterly.

6. Appendix

Symbols

ω = circular frequency of the wave
a = local sound speed
k_0 = ω/a
γ = ratio of specific heats
$\omega_g = g/a\sqrt{\gamma - 1}$ = Brunt–Väisälä frequency
$\omega_c = \gamma g/2a = a/2H$
g = acceleration due to gravity
H = pressure scale height
U = wind velocity of background field
k = propagation vector
$k^2 = k_x^2 + k_z^2$
P = static pressure
P^0 = stagnation pressure
ρ = density
u = velocity
λ_z = vertical wavelength
λ_x = horizontal wavelength
τ = wave period

REFERENCES

Bendat, J. S. and Piersol, A. G. 1966. *Measurement and analysis of random data*. Wiley, New York.

Blackman, R. B. and Tukey, J. W. 1958. *The measurement of power spectra*. Dover.

Burnside, W. 1889. On the small wave motion of a heterogeneous fluid under gravity. *Proc. London Math. Soc.* 392–397.

Eckart, C. 1960. *Hydrodynamics of oceans and atmospheres*. Pergamon, New York.

Gossard, E. E. 1962. Vertical flux of energy into the lower ionosphere from internal gravity waves generated in the troposphere. *J. Geophysical Research 67*, 745–757.

Görtler, H. 1943. Über eine Schwingungserscheinung in Flüssigkeiten mit stabiler Dichteschichtung. *Z. für Angew. Math. und Mech. 23* (2), 65–71.

Greennow, J. S. and Neufeld, E. L. 1955. *Phil. Mag. 46*, 549–561.

Greenhow, J. S. and Neufeld, E. L. 1959. *J. Geophys. Res. 64*, 2129–2133.

Hensler, L. H. 1958. *Aust. J. Phys. 11*, 79–90.

Hines, C. O. 1960. Internal atmospheric gravity waves at ionospheric heights. *Can. J. Phys. 38*, 1441–1481.

Hodges, R. R. 1969. Eddy diffusion coefficients due to instabilities in internal gravity waves. *J. Geophys. Res. 24*, 4087–4090.

Horvath, J. J., Simmons, R. W. and Brace, L. H. 1962. Theory and implementation of the pitot-static technique for upper atmospheric measurements. *Space Physics Research Lab. Scientific Report No. 1*, University of Michigan.

Jones, W. L. 1969. Ray tracing for internal gravity waves. *J. Geophys. Res. 74*, 2028–2033.

Justus, C. G. and Edwards, H. D. 1971. Winds observed from July 1968 through November 1970 in the 83 to 216 km altitude region. *Final Report, AFCRL Contract F 19628-68-C-0081*.

Lamb, H. 1909. On the theory of waves propagated vertically in the atmosphere. *Proc. London Math. Soc.* 122–141.

Lindzen, R. S. 1968. The application of classical atmospheric tidal theory. *Proc. R. Soc. A303*, 299–316.

Lindzen, R. S. 1971. Tides and gravity waves in the upper atmosphere. In *Mesospheric models and related experiments* (edited by G. Fiocco). D. Reidel, Dodrecht, Holland.

Love, E. H. 1891. Wave motion in a heterogeneous heavy liquid. *Proc. London Math. Soc.*, 307–316.

Midgley, J. E. and Liemohn, H. B. 1966. Gravity waves in a realistic atmosphere. *J. Geophys. Res. 71*, 3729–3748.

Munro, G. H. 1958. *Aust. J. Phys. 11*, 91–112.

Nordberg, W. and Smith, W. 1964. The rocket grenade experiment. *NASA Technical Note D-2107*.

Pitteway, M. L. V. and Hines, C. O. 1963. The viscous damping of atmospheric gravity waves *Can. J. Phys. 41*, 1935–1948.

Smith, W. S., Katchen, L. B. and Theon, J. S. 1968. Grenade experiments in a program of synoptic meterological measurements. *Meteorol. Monogr. 8*, (31), 170–175.

Smith, W. S., Theon, J. S., Swartz, P. C., Casey, J. F. and Horvath, J. J. 1969. Temperature, pressure, density and wind measurements in the stratosphere and mesosphere, 1967. *NASA TR-R-316*.

Smith, W. S., Theon, J. S., Wright, D. U., Casey, J. F. and Horvath, J. J. 1972. Measurements of the structure and circulation of the stratosphere and mesosphere, 1970. *NASA TR-R-391*.

Smith, W. S., Theon, J. S., Wright, D. U., Ramsdale, D. J. and Horvath, J. J. 1974. Measurements of the structure and circulation of the stratosphere and mesosphere, 1971–72. *NASA TR-R-416*.

Tolstoy, I. 1973. *Wave propagation*. McGraw-Hill.

Volland, H. 1969a. Full wave calculation of gravity wave propagation through the thermosphere. *J. Geophys. Res. 74*, 1786–1795.

Volland, H. 1969b. The upper atmosphere as a multiple refractive medium for neutral air motions. *J. Atmos. Terr. Phys. 31*, 491–514.

Witt, G. 1962. Height, structure and displacement of noctilucent clouds. *Tellus 14*, 1–18.

Yih, C.-S. 1965. *Dynamics of non-homogeneous fluids*. Macmillan, New York.

ВЕРТИКАЛЬНЫЕ ПРОФИЛИ ТЕМПЕРАТУРЫ И ПЛОТНОСТИ В АРКТИЧЕСКОЙ МЕЗОСФЕРЕ, АНАЛИЗИРУЕМЫЕ КАК ГРАВИТАЦИОННЫЕ ВОЛНЫ

С трех высокоширотных мест зимой были проведены три серии ракетных зондирований. В первой серии в течение периода 13 часов из Форта Черчилль (Канада 59° с.ш.) 31 января и 1 февраля 1967 г. были запущены четыре трубки Пито для измерения давления. Вторая серия, состоявшая из одного измерения трубкой Пито и двух гранатных акустических зондирований, была проведена в течение трехчасового периода 13–14 января 1970 г. Независимо были измерены профили ветра и температуры и один профиль плотности для получения термодинамической структуры, структуры ветра и, таким образом, их взаимозависимость в мезосфере. Третья серия измерений была проведена из Пойнт Бэрроу, Аляска (71° с.ш.) 6 декабря 1971 г. Эта серия состояла из пяти зондирований, из которых два первых и два последних зондирования состояли из измерений давления трубкой Пито и гранатных акустических зондирований. Профили температуры из всех зондирований в каждой серии были осреднены, и гладкая кривая (или серии гладких кривых) были проведены через точки. Была рассчитана гидростатическая атмосфера, основанная на осредненных измеренных профилях температуры, и отклонения от среднего строения атмосферы были проанализированы в терминах теории гравитационных волн. Вертикальная длина волн в отклонениях была 10–20 км, причем амплитуды волн медленно росли с высотой. Экспериментальные данные были согласованы с вычисленными гравитационными волнами с периодом от 15 до 80 минут и горизонтальными длинами волн от 60 до 280 км. Наша интерпретация в общем согласуется с результатами других исследователей, которые изучали гравитационно-акустические волны в атмосфере. Измерения ветра согласуются с термодинамическими измерениями. Результаты также указывают на то, что гравитационные волны распространяются с востока на запад с горизонтальной фазовой скоростью, равной приблизительно 60 м/с.

Paper 22

NASA Technical Memorandum 79630 August 1978

A VORTEX MODEL FOR TRANSPORT IN THE POLAR STRATOSPHERE

I. J. Eberstein, F. Y. Yap and V. Viers

ABSTRACT

A semi-empirical model based on a Gaussian vorticity distribution has been developed for determining eddy diffusivity and wind transport distributions in the polar stratosphere. The model uses as input data pressure surface heights measured at periods of the year when the stratospheric polar vortex exhibits nearly circular patterns around the pole. The components of the polar wind velocities that result from a Prandtl eddy viscosity distribution are found to be in general agreement with those obtained by other investigators.

Paper 23

Reprinted from *Journal of Geophysical Research*, 82, pp. 953-988, 1978.

Ozone Profiles and Chemical Loss Rates in the Tropical Stratosphere Deduced From Backscatter Ultraviolet Measurements

J. E. FREDERICK,[1] B. W. GUENTHER,[2] P. B. HAYS,[1] AND D. F. HEATH[2]

Analysis of data obtained by the backscatter ultraviolet (BUV) experiment on the Atmosphere Explorer E satellite has provided equatorial ozone mixing ratio profiles for equinox and solstice conditions. The combination of these results with a pure oxygen chemical model yields the rate of odd oxygen loss due to the sum of the odd hydrogen, nitrogen, and chlorine cycles. Use of recent mid-latitude stratospheric measurements of HO_x, NO_x, and ClO_x with the BUV data provides an independent calculation of the catalytic loss. Below 45 km the agreement between the two sets of loss rates is satisfactory. At higher altitudes the odd hydrogen cycle provides far more O_x loss than can be tolerated by the BUV measurements if the photodissociation of O_2 is the only source and has the currently accepted magnitude. The results suggest either a tropical HO_x concentration smaller than is now believed or the presence of a very large source of odd oxygen in the upper stratosphere and lower mesosphere.

Introduction

Satellite measurements of the backscattered solar radiation which emerges from the earth's atmosphere allow near-continuous monitoring of the vertical profile of ozone between approximately 30 and 55 km. Analysis of early backscatter ultraviolet (BUV) data from the Ogo 4 and Nimbus 4 satellites has provided a global mapping of the latitudinal and seasonal distribution of upper stratospheric ozone [*London et al.*, 1977; *Krueger et al.*, 1973] on a spacial grid much finer than is possible with the ground-based Umkehr method. In this paper we present tropical ozone profiles obtained by the BUV instrument on the Atmosphere Explorer E spacecraft (AE-E) and compare these to earlier satellite and rocket measurements at low and middle latitudes.

The history of stratospheric photochemistry shows the development of increasingly sophisticated models, from the pure oxygen atmosphere of *Chapman* [1930] to the inclusion of hydrogen [*Bates and Nicolet*, 1950; *Hunt*, 1966], nitrogen [*Crutzen*, 1970, 1971], and more recently, chlorine [*Molina and Roland*, 1974; *Stolarski and Cicerone*, 1974]. There is no a priori reason to believe that the current theories are complete. Indeed, comparison of the predictions of photochemical theories with measurements in the upper stratosphere and mesosphere indicates that current models underestimate the true ozone abundance [*Sze*, 1977; *Frederick et al.*, 1977; *Liu and Cicerone*, 1977]. In an attempt to shed light on this question, we combine the BUV data with recent measurements of upper stratospheric NO_x, ClO_x, and HO_x. In particular, we wish to determine if the catalytic loss of odd oxygen required by the BUV data is in reasonable agreement with that supplied by the radical measurements.

Experiment Description and Data Availability

The BUV instrument on AE-E is essentially the same as those instruments flown on the Nimbus satellites and has been discussed by *Heath et al.* [1973]. The spectrophotometer monitors the vertically emergent intensity from the atmosphere at 12 wavelengths between 2550 Å and 3400 Å with a 10-Å resolution. Only those wavelengths for which single molecular scattering is a good approximation, $\lambda < 3000$ Å, enter into the ozone profile determination. *Mateer* [1972] has discussed the single-scattering formulation of the radiative transfer equation and the methods used to infer the O_3 distribution.

The near-equatorial orbit of AE-E (inclination ≈ 20°) is well suited to studies aimed at testing photochemical theories of the stratosphere. The small average solar zenith angle allows maximum penetration of the solar ultraviolet flux, making photochemical processes more competitive with transport in the middle stratosphere than they are at higher latitudes. In addition, seasonal variations in density and temperature are rather small near the equator. These two advantages of the low inclination orbit allow one to isolate chemical loss from transport and seasonal effects with reasonable certainty. Finally, an orbit at near-constant latitude provides for the accumulation of a large data base over a limited geographic region so that the true mean state of the upper stratospheric ozone profile may be assessed.

Data presented in this report are taken from two time periods, near December solstice (1975) and March equinox (1976). The equinox data set contained 4 orbits with a total of 89 ozone profiles, while the solstice analysis included 18 orbits with a total of 1081 profiles. Additional data for the time periods of interest were not available. Figure 1 presents the average ozone volume mixing ratios for solstice and equinox. All profiles measured in the appropriate time interval were included in the averages, and the half-width of the error bars on the solstice result denotes 1 standard deviation from the mean. For easy reference, Table 1 presents the average volume mixing ratios and the standard deviations at six pressure levels. Although some differences between the solstice and equinox results appear, the disparity between the sizes of the two data sets precludes definite statements concerning variations.

Comparison to Previous Results

Few measurements of the ozone profile above 30 km in the tropics are available. Figure 2 compares the AE-E solstice profile with (1) the equatorial stellar occultation measurements of *Riegler et al.* [1977] from the Copernicus satellite, (2) the backscatter ultraviolet results of *London et al.* [1977] from Ogo 4 averaged over the latitude band 15°N–15°S for January and February 1968, (3) the *Krueger and Minzner* [1976] mid-latitude model, and (4) a tropical model based on the Krueger-Minzner composite. The final profile was derived by estimating the ozone gradient between tropical and middle latitudes from the results of *London et al.* [1977] and applying this

[1] Space Physics Research Laboratory, Department of Atmospheric and Oceanic Science, University of Michigan, Ann Arbor, Michigan 48109.

[2] Atmospheric and Hydrospheric Applications Division, Goddard Space Flight Center, Greenbelt, Maryland 20771.

Paper number 7C0796
0148-0227/78/027C-0796$03.00

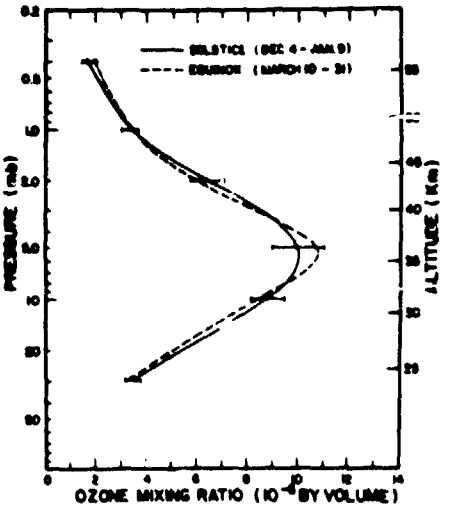


Fig. 1 Average ozone volume mixing ratio profiles for solstice and equinox Error bars denote plus and minus 1 standar ' deviation of the mean.

correction to the Krueger-Minzner model. The U S. Standard Atmosphere Supplements (1966) 15°N annual model has been used to convert mixing ratios to number densities.

The AE-E values are larger than those predicted by Ogo 4 above 30 km. The ratio of the AE-E values to the Ogo 4 values varies from 1.1 to 1 3 above 35 km. Above 40 km the AE-E results agree better with the Krueger-Minzner mid-latitude model than with the derived tropical model We stress that this good agreement is probably fortuitous, being the result of systematic calibration differences, since the results of *London et al.* [1977] show significant latitudinal gradien . below 30 km the single-scattering approximation used in the BUV inversion begins to break down, so the crossover of the AE-E and Krueger-Minzner models is not significant The *Riegler et al* [1977] values lie a factor of 2-3 above any of the other results The Copernicus profile refers to nighttime tropical conditions, while all others were measured in daylight Current chemical models predict little or no diurnal variation in O_3 between 50 and 55 km. If we accept the differences shown in Figure 2 as real, we must admit a serious deficiency in our present understanding of stratospheric photochemistry As will be shown later, there is additional evidence that such failings exist. We note, however, that satellite-borne infrared measurements during both day and night show little or no diurnal variation near 50 km in the tropics [*Gille et al*, 1977] In view of this we find it difficult to accept the Copernicus results.

Chemical Model and Odd Oxygen Loss Rates

Table 2 lists the reactions and their rate coefficients employed in the following analysis. The total odd oxygen concentration, $[O_x] = [O(^1D)] + [O(^3P)] + [O_3]$, is given by the continuity equation

$$\frac{\partial [O_x]}{\partial t} + \frac{\partial \phi(O_x)}{\partial z} = P(O_x) - L(O_x) - L(\text{cat}) \qquad (1)$$

where

- $P(O_x)$ odd oxygen production rate, cm^{-3} s^{-1} (reactions (R1) and (R2) of Table 2);
- $L(O_x)$ odd oxygen loss rate involving reactions among $O(^1D)$, $O(^3P)$, and O_3 only (reactions (R6), (R7), and (R8));
- $L(\text{cat})$ odd oxygen loss rate due to catalytic cycles of NO_x, ClO_x, and HO_x (reactions (R3), (R4), (R5), (R13), (R14), (R16), (R17), (R18), (R19), and (R20));
- $\phi(O_x)$ vertical flux of odd oxygen, cm^{-2} s^{-1}, parameterized by an eddy diffusion coefficient, $\phi = -K[M]\,\partial/\partial z([O_x]/[M])$.

We denote the process of averaging over a diurnal cycle by a bar placed over the quantity of interest and assume that $\partial[\overline{O_x}]/\partial t = 0$ to obtain

$$\overline{L(\text{cat})} = \overline{P(O_x)} - \overline{L(O_x)} - \frac{\partial \overline{\phi(O_x)}}{\partial z} \qquad (2)$$

Use of a tropical model atmosphere for $[O_x]$, $[M]$, and temperature in conjunction with the BUV ozone profiles and the chemistry of a pure oxygen atmosphere allows an evaluation of the right-hand side of (2) To mimic vertical transport in the tropics, we use the large eddy diffusion coefficient of *Schmeltekopf et al* [1977] at 25 km, $K = 7.5 \times 10^4$ cm^2 s^{-1}, which is based on N_2O measurements from Panama (5.5°N) We then extend the K profile linearly upward to a value of 1×10^5 cm^2 s^{-1} at 40 km However, above 35 km the transport term in (2) is not significant We then know the catalytic loss rate required to reproduce the BUV measurements. Use of available NO_x, ClO_x, and HO_x profiles with the BUV results and the reactions of Table 2 gives a second evaluation of $\overline{L(\text{cat})}$ which is independent of the source of odd oxygen on which the previous result largely depends In performing these calculations we assume that the reactions which interconvert odd constituents proceed much faster than the sources and sinks Thus the concentrations of all odd constituents in a given catalytic cycle can be computed easily when a measurement of only one of these is available For example, given the measured O_3 values, $[O(^1D)]$ and $[O(^3P)]$ are readily deduced from reactions (R9)-(R12) The process of diurnal averaging is particularly simple in this case, since the O_3 concentration is constant over a 24-hour period at most of the altitudes considered here Recall, however, the discussion of the previous section

Figure 3 compares the $O(^3P)$ concentration computed from the BUV data at a solar zenith angle of 50° with the measurement of *Anderson* [1975] at 32°N latitude The agreement is

TABLE 1 Average Ozone Volume Mixing Ratios and Standard Deviations for Solstice and Equinox

	Solstice		Equinox	
P, mbar		Standard Deviation	Mixing Ratio	Standard Deviation
0.4	1.68×10^{-6}	2.20×10^{-7}	1.98×10^{-6}	3.22×10^{-7}
1.0	3.32×10^{-6}	3.39×10^{-7}	3.37×10^{-6}	2.14×10^{-7}
2.0	6.41×10^{-6}	6.65×10^{-7}	5.96×10^{-6}	4.80×10^{-7}
5.0	1.00×10^{-5}	1.02×10^{-6}	1.08×10^{-5}	8.55×10^{-7}
10.0	8.78×10^{-6}	6.76×10^{-7}	8.38×10^{-6}	1.09×10^{-6}
30.0	3.45×10^{-6}	2.88×10^{-7}	3.16×10^{-6}	5.41×10^{-7}

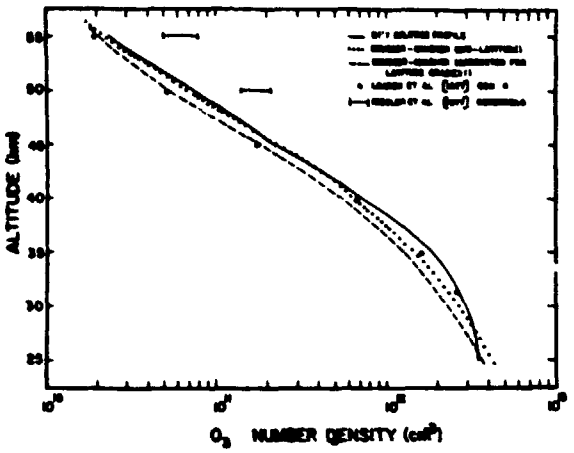


Fig 2 Comparison of the solstice BUV ozone profile with other tropical and mid-latitude results

quite good between 35 and 40 km; however, at lower altitudes the calculation falls below the measurement. Multiple scattering effects and a surface albedo of 0.35 were included in the calculation from *Luther and Gelinas* [1976]. Below 35 km, meteorological variability in ozone increases as the re[illegible] time becomes long This increase, combined with uncer[illegible] in the albedo corrections at the lowest altitudes, contributes to the discrepancies in Figure 3. The favorable comparison above 35 km supports the validity of the chemical reaction set. If we use the latitudinal gradients observed by *London et al* [1977] to refer the AE-E results to 30°N, the computed $O(^3P)$ increases by a factor of 1.2–1.3 below 40 km.

Figure 4 presents the diurnally averaged profiles of $\bar{P}(O_x)$, $\bar{L}(O_x)$, and $\bar{L}$(cat) for the solstice ozone result The need for a substantial catalytic loss is obvious. For computing the individual catalytic losses due to the nitrogen, chlorine, and hydrogen cycles, $\bar{L}(NO_x)$, $\bar{L}(ClO_x)$, and $\bar{L}(HO_x)$, respectively, we have adopted mid-latitude measurements in the absence of tropical data. We employ two models each of odd nitrogen ($[NO_x] = [NO] + [NO_2]$) and odd chlorine ($[ClO_x] = [Cl] + [ClO]$) to represent the likely range of natural variations. We refer to these as the 'large' and 'small' NO_x and ClO_x profiles. No single measurement of NO_x reported to date has spanned the entire altitude region of interest, 25–55 km. To obtain the large NO_x model, we combine the unpublished NO data of J. J. Horvath between 42 and 55 km with the total NO_x measurements of *Ackerman et al.* [1975] between 25 and 36 km. The former data were obtained by a chemiluminescent rocket probe in the manner described by *Mason and Horvath* [1976] and, at the altitudes of the measurement, are a good representation of total NO_x. The original values of *Mason and Horvath* [1976] are now believed to underestimate the true concentration due to uncertainties in the instrument background signal (J. J. Horvath, personal communication, 1976). The adopted high-altitude results merged rather smoothly with the total NO_x measurements of *Ackerman et al.* [1975], which were obtained by balloon-borne solar infrared absorption methods near sunset. Although NO and NO_2 individually vary with time at high solar zenith angles, their sum remains constant. However, addition of the two constituents measured separately leads to large error bars. The morning NO results of *Drummond et al.* [1977] from a balloon-borne chemiluminescent probe in Wyoming form the basis of the small NO_x model. Nitrogen dioxide concentrations were computed from the measurements by using a solar zenith angle of 76° and the small ClO model discussed below. The results were extrapolated to 55 km from their peak altitude of 45 km. Table 3 lists the final large and small NO_x models.

The availability of ClO_x data is even less satisfactory than that of NO_x. We have adopted large and small ClO models based on the balloon-borne resonance fluorescence measurements of *Anderson and Margitan* [1977] conducted from P[illegible]stine, Texas, on October 2 and December 8, 1976, respectively. Atomic chlorine concentrations were then computed from the reactions of Table 2, neglecting the small effect of ClO photo-

TABLE 2. Reactions and Rate Coefficients Used in the Data Analysis

No	Reaction	Rate	Reference
(R1)	$O_2 + h\nu \rightarrow O(^3P) + O(^3P)$		*Ackerman* [1971]
(R2)	$O_2 + h\nu \rightarrow O(^3P) + O(^1D)$	$J_1 + J_2 = 4.3 \times 10^{-10}$*	*Park* [1974]
(R3)	$NO + O_3 \rightarrow NO_2 + O_2$	$k_3 = 9.0 \times 10^{-12} \exp(-1300/T)$	*Bortner et al.* [1973]
(R4)	$NO_2 + O(^3P) \rightarrow NO + O_2$	$k_4 = 9.1 \times 10^{-12}$	*Bortner et al* [1973]
(R5)	$NO_2 + h\nu \rightarrow NO + O(^3P)$	$J_5 = 4.8 \times 10^{-3}$*	*Brasseur and Nicolet* [1973]
(R6)	$O(^3P) + O_2 \rightarrow O_3 + O_2$	$k_6 = 1.1 \times 10^{-11} \exp(-2150/T)$	*Nicolet* [1975]
(R7)	$O(^1D) + O_2 \rightarrow O_3 + O_2$	$k_7 = 2.5 \times 10^{-10}$	*Bortner et al* [1973]
(R8)	$O(^3P) + O(^3P) + M \rightarrow O_2 + M$	$k_8 = 4.7 \times 10^{-33}(300/T)^2$	*Nicolet* [1975]
(R9)	$O(^3P) + O_2 + M \rightarrow O_3 + M$	$k_9 = 1.1 \times 10^{-34} \exp(510/T)$	*Nicolet* [1975]
(R10)	$O_3 + h\nu \rightarrow O_2 + O(^3P)$	$J_{10} = 2.4 \times 10^{-4}$	*Ackerman* [1971]
(R11)	$O_3 + h\nu \rightarrow O_2 + O(^1D)$	$J_{11} = 2.6 \times 10^{-3}$	*Ackerman* [1971]
(R12)	$O(^1D) + M \rightarrow O(^3P) + M$	$k_{12} = 3.2 \times 10^{-11}$	*Streit et al.* [1976]
(R13)	$Cl + O_3 \rightarrow ClO + O_2$	$k_{13} = 3.6 \times 10^{-11} \exp(-318/T)$	*Liu et al.* [1976]
(R14)	$ClO + O(^3P) \rightarrow Cl + O_2$	$k_{14} = 1.2 \times 10^{-10} \exp(-250/T)$	*Liu et al.* [1976]
(R15)	$ClO + NO \rightarrow Cl + NO_2$	$k_{15} = 1.7 \times 10^{-11}$	*Liu et al* [1976]
(R16)	$OH + O(^3P) \rightarrow H + O_2$	$k_{16} = 4 \times 10^{-11}$	*Nicolet* [1975]
(R17)	$H + O_3 \rightarrow OH + O_2$	$k_{17} = 1.2 \times 10^{-10} \exp(-560/T)$	*Nicolet* [1975]
(R18)	$HO_2 + O(^3P) \rightarrow OH + O_2$	$k_{18} = 8 \times 10^{-11} \exp(-500/T)$	*Nicolet* [1975]
(R19)	$OH + O_3 \rightarrow HO_2 + O_2$	$k_{19} = 1.3 \times 10^{-12} \exp(-950/T)$	*Nicolet* [1975]
(R20)	$HO_2 + O_3 \rightarrow OH + O_2 + O_2$	$k_{20} = 1 \times 10^{-14} \exp(-1250/T)$	*Nicolet* [1975]
(R21)	$H + O_2 + M \rightarrow HO_2 + M$	$k_{21} = 6 \times 10^{-32} \exp(290/T)$	*Nicolet* [1975]

* All dissociation rates listed refer to an altitude of 50 km for solstice conditions and have been diurnally averaged. Reported values assume that only the direct solar beam is present The rates were corrected for multiple scattering and a surface albedo of 0.35 by using the results of *Luther and Gelinas* [1976] before use in calculations reported in the text

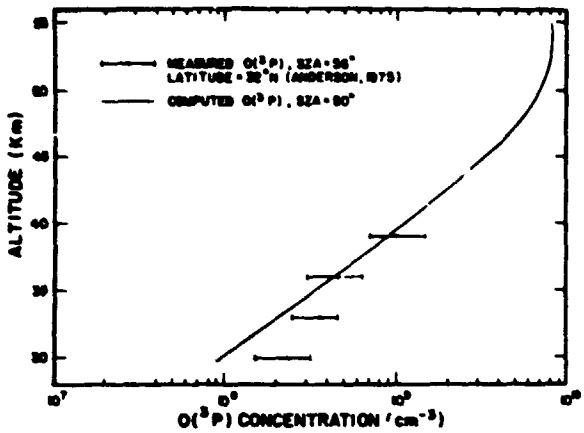


Fig. 3. Comparison of the $O(^3P)$ measurement of *Anderson* [1975] with the profile deduced from the solstice ozone measurements.

dissociation. Table 3 presents the adopted large and small ClO_x models.

The total odd hydrogen concentration ($[HO_x] = [H] + [OH] + [HO_2]$) is not constant during a 24-hour period as was the case with NO_x. A measurement at a specific local time combined with the interconversion reactions of HO_x does not specify the odd hydrogen concentration at all other times. We therefore adopt theoretical calculations from the time-dependent model of S. C. Liu (personal communication, 1977) by using rate coefficients such that the mid-latitude OH measurement of *Anderson* [1976] is reproduced at a solar zenith angle of 80°. *Liu et al.* [1976] have described similar calculations using a diurnally averaged model. Figure 5 presents the final OH profiles used at various solar zenith angles and the *Anderson* [1976] measurement. We consider the limitations of using mid-latitude radical concentrations with the tropical ozone data in a later section.

The instantaneous O_x catalytic loss rate derived from the reactions of Table 2 is

$$L'(\text{cat}) = L(NO_x) + L(ClO_x) + L(HO_x) \qquad (3)$$

TABLE 3 NO_x and ClO_x Models Used in the Catalytic Loss Rate Calculations

Altitude, km	Large NO_x,* cm^{-3}	Small NO_x,† cm^{-3}	Large ClO_x,‡ cm^{-3}	Small ClO_x,‡ cm^{-3}
55	9.80×10^7	9.12×10^7		
50	1.87×10^8	1.72×10^8	2.40×10^8	6.00×10^8
45	4.50×10^8	3.20×10^8	1.60×10^7	4.50×10^8
40	1.18×10^9	6.12×10^8	9.20×10^7	2.40×10^8
35	2.10×10^9	9.40×10^8	4.70×10^8	9.00×10^8
30	3.00×10^9	1.30×10^9	6.70×10^8	6.40×10^7
25	4.00×10^9	1.80×10^9	4.20×10^8	4.00×10^7

* Based on unpublished NO measurements by J. J. Horvath above 42 km and data of *Ackerman et al.* [1975] at lower altitudes. ($[NO_x] = [NO] + [NO_2]$)

† Based on NO measurement of *Drummond et al.* [1977] up to 45 km and extrapolated to higher altitudes.

‡ Based on data of *Anderson and Margitan* [1977] up to 40 km and extrapolated to higher altitudes.

where

$$L(NO_x) = k_3[O_3][NO] + k_4[O(^3P)][NO_2] - J_5[NO_2] \qquad (4)$$

$$L(ClO_x) = k_{13}[O_3][Cl] + k_{14}[O(^3P)][ClO]$$

$$L(HO_x) = 2k_{16}[O(^3P)][OH] + 2k_{18}[O_3][OH] \qquad (5)$$

$$\frac{[NO_2]}{[NO]} = \frac{k_3[O_3] + k_{13}[ClO]}{k_4[O(^3P)] + J_5} \qquad (6)$$

and

$$\frac{[Cl]}{[ClO]} = \frac{k_{14}[O(^3P)] + k_{15}[NO]}{k_{13}[O_3]} \qquad (7)$$

We add a prime to the loss rate computed in (3) to distinguish it from the results of (2) and Figure 4. The time constant for chemical removal of odd oxygen in the upper stratosphere is longer than the duration of day and night so that the measured ozone concentrations represent a response to a diurnally averaged solar flux. For comparison to Figure 4 the predictions of (3) must be diurnally averaged. We do this by computing $L'(\text{cat})$ at 1-hour intervals during the course of the day and numerically averaging the results over 24 hours. The total NO_x and ClO_x concentrations were held constant, but the ratios of the individual odd species were allowed to vary during the day

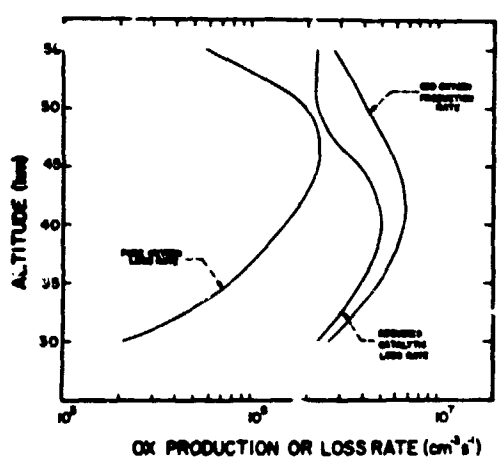


Fig. 4. Diurnally averaged profiles of the odd oxygen production rate, $\bar{P}(O_x)$, the loss rate in a pure oxygen atmosphere, $\bar{L}(O_x)$, and the catalytic loss rate $\bar{L}(\text{cat})$ required to reproduce the BUV solstice ozone profile

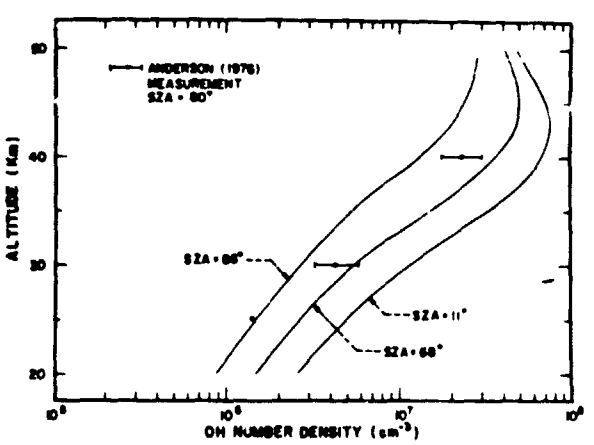


Fig. 5. Comparison of hydroxyl radical calculations for various solar zenith angles (S. C. Liu, personal communication, 1977) with the measurement of *Anderson* [1976]

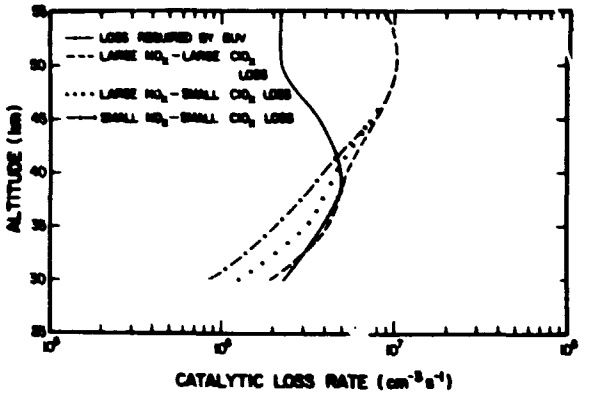


Fig 6 Diurnally averaged catalytic loss rates computed by using NO_x, ClO_x, and HO_x models and the rate required to reproduce the solstice ozone profile The HO_x catalytic cycle becomes the dominant loss process above 40 km.

The averaging procedure was done for the standard HO_x model and four combinations of the NO_x and ClO_x profiles, large NO_x-large ClO_x, large NO_x-small ClO_x, small NO_x-large ClO_x, and small NO_x-small ClO_x.

Comparison of Catalytic Loss Rates

Figure 6 presents the diurnally averaged catalytic loss rates computed from the adopted OH profile and three of the NO_x-ClO_x combinations. Also shown is the $\bar{L}$(cat) result required by the solstice ozone profile. Between 30 and 42 km the large NO_x-large ClO_x model approximately provides the loss rate demanded by the BUV measurements. The presence of latitudinal gradients, meteorological variability, and uncertainties in the radical measurements requires that one not demand strict agreement If we assume latitudinal gradients in NO_x and ClO_x similar to those in O_3, then the tropical profiles will be smaller than those used here by a factor of 1.1–1.3 Model calculations for NO_x by *Prinn et al.* [1975] show this trend; however, the measurements of *Schmeltekopf et al* [1977] reveal N_2O concentrations in the tropics near 30 km which are a factor of 2 above the theoretical results In view of the uncertainty in stratospheric transport parameters above 30 km, it is not clear that motions can reverse the latitudinal gradient in NO_x established by the large tropical source. Similar arguments apply to ClO_x From the available information we can conclude only that the ozone profiles measured by AE-E are consistent with large but reasonable concentrations of NO_x and ClO_x.

Above 42 km the O_x catalytic loss due to HO_x becomes important, and a significant discrepancy appears in Figure 6. Hydroxyl radical concentrations consistent with the *Anderson* [1976] measurement at 32°N in January provide far more loss than can be tolerated by the BUV results. At 45 km the discrepancy is a factor of 2.1 and increases to a factor of 4.7 at 50 km. At these altitudes the O_x loss rate in a pure oxygen atmosphere is ½ or less of the production rate so that the $\bar{L}$(cat) prediction of (2) cannot be severely in error. *Barnett* [1976] has observed large temporal variations in the total OH column abundance. The profile of *Anderson* [1976] implies a column abundance typical of those of Barnett. However, early rocket-borne measurements of OH between 45 and 70 km by *Anderson* [1971] at a solar zenith angle of 86° gave results much smaller than those used here If the time-dependent OH calculations are normalized to the older results, one obtains satisfactory agreement between the two sets of O_x loss rates. However, present knowledge of the reaction rates involved in the OH calculations [*Chang and Kaufman*, 1977] supports the more recent, large measurements. The calculations of *Prinn et al.* [1975] show minor latitudinal variations in OH between 32°N and the equator in winter. The gradient which does exist implies more OH in the tropics than at mid-latitudes, where the measurement used here applies. We must conclude that either the average OH concentration within 20° of the equator is much smaller than that predicted by current photochemical models and measured at 32°N or the odd oxygen catalytic loss rate is larger than that predicted by (2).

Conclusions

The tropical ozone profiles obtained by the AE-E BUV experiment are in general agreement with other low-latitude data. The large quantity of information shows that temporal variations in tropical ozone are small, at least over time scales of the order of 3 months.

At present only a small amount of data exists on which to base calculations of the O_x catalytic loss rate. The largest available NO_x and ClO_x data are able to supply the loss required by the BUV results between 30 and 42 km. Above 42 km the discrepancy between the O_x loss supplied by OH models and the loss deduced from the BUV data appears larger than any likely error in the input data. If we accept the OH concentration used here as representative of the mean state of the atmosphere, we then require an odd oxygen production rate larger than that supplied by current calculations of the photodissociation rate of O_2. This implies either additional sources of O_x or a very large value of $J_1 + J_2$ above 40 km.

The large variations which exist in mid-latitude values of $[NO_x]$, $[ClO_x]$, and $[HO_x]$ are responsible for major uncertainties when single measurements are compared with model calculations. The small variability in tropical ozone suggests similar behavior for other constituents. A series of NO_x, ClO_x, and HO_x measurements at low latitudes combined with ozone data could provide a better test of photochemical theories than is possible with results presently available

Acknowledgments The authors are grateful to C L Mateer for providing details on the inversion technique and to J J Horvath for giving us access to his unpublished NO measurements J G Anderson, S C Liu, J London, and R S Stolarski provided many useful comments on the original manuscript This research was supported by the National Aeronautics and Space Administration under contract NAS 5-23006

References

Ackerman, M , Ultraviolet solar radiation related to mesospheric processes, in *Mesospheric Models and Related Experiments*, edited by G Fiocco, p. 189, D Reidel, Hingham, Mass., 1971.

Ackerman, M , J C Fontanella, D Frimout, A Girard, M Louisnard, and C Müller, Simultaneous measurements of NO and NO_2 in the stratosphere, *Planet Space Sci.* *23*, 651, 1975

Anderson, J. G., Rocket measurement of OH in the mesosphere, *J Geophys Res* , *76*, 7820, 1971.

Anderson, J. G., The absolute concentration of $O(^3P)$ in the earth's stratosphere, *Geophys Res Lett* , *2*, 231, 1975.

Anderson, J G , The absolute concentration of $OH(X^2\pi)$, in the earth's stratosphere, *Geophys Res Lett* , *3*, 165, 1976

Anderson, J G , and J J Margitan, Atoms and diatomic radicals in the earth's stratosphere The in situ observation of $O(^3P)$, $OH(X^2\pi)$, $Cl(^2P)$ and ClO (abstract), *Eos Trans AGU* , *58*, 463, 1977

Barnett, C R , Terrestrial OH abundance measurement by spectroscopic observation of resonance absorption of sunlight, *Geophys Res Lett* *3*, 319, 1976

Bates D R , and M Nicolet, The photochemistry of atmospheric water vapor, *J Geophys Res* *55*, 301, 1950

Bortner, M. H., R. H. Kummler, and T. Baurer, Summary of suggested rate constants, in *Defense Nuclear Agency Reaction Rate Handbook*, revision 3, chap. 24, DOD Information and Analysis Center, Santa Barbara, Calif., 1973.

Brasseur, G., and M. Nicolet, Chemospheric processes of nitric oxide in the mesosphere and stratosphere, *Planet Space Sci.*, *21*, 939, 1973.

Chang, J. S., and F. Kaufman, Determination of an upper limit to the rate constant of the reaction $OH + HO_2 \rightarrow H_2O + O_2$ (abstract), *Eos Trans. AGU*, *58*, 464, 1977.

Chapman, S., A theory of upper atmospheric ozone, *Mem Roy Meteorol Soc.*, *3*, 103, 1930.

Crutzen, P. J., The influence of nitrogen oxides on the atmospheric ozone content, *Quart J. Roy. Meteorol Soc.*, *96*, 320, 1970.

Crutzen, P. J., Ozone production rates in an oxygen-hydrogen-nitrogen oxide atmosphere, *J. Geophys Res.*, *76*, 7311, 1971.

Drummond, J. W., J. M. Rosen, and D. G. Hofmann, Balloon-borne chemiluminescent measurement of NO to 45 km, *Nature*, *265*, 319, 1977.

Frederick, J. E., P. B. Hays, and S. C. Liu, The need for a revised theory of mesospheric ozone (abstract), *Eos Trans. AGU*, *58*, 462, 1977.

Gille, J. C., P. L. Bailey, G. P. Anderson, and S. Nolte, Comparison between limb radiance inversion radiometer retrievals and in situ measurements (abstract), *Eos Trans AGU*, *58*, 461, 1977.

Heath, D. F., C. L. Mateer, and A. J. Krueger, The Nimbus-4 backscatter ultraviolet (BUV) atmospheric ozone experiment Two years operation, *Pure Appl Geophys.*, *106–108*, 1278, 1973.

Hunt, B. G., Photochemistry of ozone in a moist, atmosphere, *J Geophys Res.*, *71*, 1385, 1966.

Krueger, A. J., and R. A. Minzner, A mid-latitude ozone model for the 1976 U.S. Standard Atmosphere, *J Geophys Res.*, *81*, 4477, 1976.

Krueger, A. J., D. F. Heath, and C. L. Mateer, Variations in the stratospheric ozone field inferred from Nimbus satellite observations, *Pure Appl. Geophys.*, *106–108*, 1254, 1973.

Liu, S. C., and R. J. Cicerone, Comparison of theory and measurements of ozone above 35 km (abstract), *Eos Trans AGU*, *58*, 462, 1977.

Liu, S. C., T. M. Donahue, R. J. Cicerone, and W. L. Chameides, Effect of water vapor on the destruction of ozone in the stratosphere perturbed by ClX or NO_x pollutants, *J Geophys Res.*, *81*, 3111, 1976.

London, J., J. E. Frederick, and G. P. Anderson, Satellite observations of the global distribution of stratospheric ozone, *J Geophys Res.*, *82*, 2543, 1977.

Luther, F. M., and R. J. Gelinas, Effect of multiple scattering and ground reflection on atmospheric photodissociation rates, *J Geophys Res.*, *81*, 1125, 1976.

Mason, C. J., and J. J. Horvath, The direct measurement of nitric oxide concentration in the upper atmosphere by a rocket-borne chemiluminescent detector, *Geophys Res Lett.*, *3*, 391, 1976.

Mateer, C. L., Mathematics of profile inversion, proceedings of a workshop held at Ames Research Center, Moffett Field, California, July 12–16, 1971, *NASA TM X-62, 150*, pp. 1–2, Aug. 1972.

Molina, M. J., and F. S. Rowland, Stratospheric sink for chlorofluoromethanes—Chlorine atom catalyzed destruction of ozone, *Nature*, *249*, 810, 1974.

Nicolet, M., Stratospheric ozone An introduction to its study, *Rev Geophys Space Phys.*, *13*, 593, 1975.

Park, J. H., The equivalent mean absorption cross sections for the O_2 Schumann-Runge bands Application to the H_2O and NO photodissociation rates, *J Atmos Sci.*, *31*, 1893, 1974.

Prinn, R. G., F. N. Alyea, and D. M. Cunnold, Stratospheric distributions of odd nitrogen and odd hydrogen in a two-dimensional model, *J Geophys Res.*, *80*, 4997, 1975.

Riegler, G. R., S. K. Atreya, T. M. Donahue, S. C. Liu, B. Wasser and J. F. Drake, UV stellar occultation measurements of nighttime equatorial ozone, *Geophys Res Lett.*, *4*, 145, 1977.

Schmeltekopf, A. L., D. L. Albritton, P. J. Crutzen, P. D. Goldan, W. J. Harrop, W. R. Henderson, J. R. McAfee, M. McFarland, H. I. Schiff, T. L. Thompson, D. J. Hoffmann, and N. T. Kjome, Stratospheric nitrous oxide altitude profiles at various latitudes, *J Atmos Sci.*, *34*, 729, 1977.

Stolarski, R. S., and R. J. Cicerone, Stratospheric chlorine A possible sink for ozone, *Can J Chem.*, *52*, 1610, 1974.

Streit, G. E., C. J. Howard, D. A. Jennings, and A. L. Schmeltekopf, Absolute rate constant determination for the deactivation of $O(^1D)$ by time resolved decay of $O(^1D) \rightarrow O(^3P)$ emission, *J Chem Phys.*, *64*, 57, 1976.

Sze, N. D., Atmospheric ozone Comparison between theory and observation (abstract), *Eos Trans AGU*, *58*, 461, 1977.

(Received February 25, 1977,
revised July 29, 1977,
accepted August 18, 1977.)

Paper 24

COUPLING PROCESSES RELATED TO THE SUN-WEATHER PROBLEM*

Richard A. Goldberg
Laboratory for Atmospheric Sciences
Goddard Space Flight Center
Greenbelt, Maryland 20771 U.S.A.

and

John R. Herman
Radio Sciences Company
624 Tulane Avenue
Melbourne, Florida 32901 U.S.A.

Physical mechanisms for coupling the energetics of solar activity to meteorological responses are reviewed. Although several hypotheses have been advanced, none can be said to be sufficiently complete to be applied to weather or climate prediction. Solar activity indicators potentially useful for forecasting are identified, including sunspots, solar flares, and magnetic sector boundary crossings. Additional experiments, studies and analyses are required before sun-weather concepts can be utilized for predicting meteorological responses.

*Invited Review for the International Solar-Terrestrial Predictions Workshop, Boulder, Colorado, April 23-27, 1979.

N81-21689

Paper 25

From the International Union of Geodesy and Geophysics (IUGG) Proceedings – Middle Atmosphere Symposium, December, 1979, Canberra, Australia.

AURORAL X-RAY DETECTION FROM ROCKET OVERFLIGHTS

R. A. Goldberg (NASA/Goddard Space Flight Center, Laboratory for Planetary Atmospheres, Greenbelt, Maryland 20771 USA)

J. R. Barcus (Dept. of Physics, Denver, Denver, Colorado 80210 and NAS-NRC Senior Resident Associate, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771 USA)

I. H. Gesell (Computer Sciences Corporation, Silver Spring, Maryland 20910 USA)

R. R. Vondrak (Radio Physics Laboratory, SRI International, Menlo Park, California 94025 USA)

Remote sensing of auroral x-rays from the topside can provide both local and global information concerning x-ray and energetic particle deposition within the middle atmosphere. Recently, satellite [1,2] and rocket [3] results have illustrated the feasibility of this technique. Usually, contamination of x-ray detectors by corpuscular radiation can severely affect the scope and accuracy of the measurement. Here we report preliminary results obtained with a rocket-borne instrument designed to operate cleanly in a precipitating particle environment. Two dimensional images of the atmospheric bremsstrahlung x-ray sources were also constructed from the detector scan produced by payload translation and coning.

VOL. 84, NO. C2 JOURNAL OF GEOPHYSICAL RESEARCH FEBRUARY 20, 1979

Sensitivity of the Predicted CO-OH-CH_4 Perturbation to Tropospheric NO_x Concentrations

SULTAN HAMEED AND JOSEPH P. PINTO

Laboratory for Planetary Atmospheres Research, Department of Mechanical Engineering State University of New York, Stony Brook, New York 11794

RICHARD W. STEWART

Goddard Space Flight Center NASA, Greenbelt, Maryland 20771

Measurements indicate that ambient NO_x ($NO + NO_2$) concentrations vary over at least an order of magnitude in the troposphere. Thus it is of interest to study atmospheric chemistry with NO_x concentrations covering this range. In this paper we present steady state calculations which show that as NO_x concentrations are increased from very low values, the hydroxyl radical concentration [OH] increases at first and then decreases. This change of behavior occurs at $[NO_2] \sim 0.23$ ppb in our model. Also, the response of the atmosphere to increased CO fluxes undergoes a qualitative change as a function of $[NO_x]$. At low levels of $[NO_x]$, an increase in CO fluxes into the atmosphere depletes [OH] and increases $[CH_4]$ and $[H_2]$, as reported in the literature. However, at high values of $[NO_x]$, increased CO fluxes give rise to an increase in [OH] and, consequently, a decrease in $[CH_4]$ and $[H_2]$.

1. INTRODUCTION

Anthropogenic carbon monoxide emissions constitute a significant fraction of the tropospheric CO source [*Jaffe*, 1975] and therefore are important in determining ambient CO concentrations. Since carbon monoxide reacts with the hydroxyl radical, OH, which in turn plays a central role in tropospheric photochemistry, any increase in anthropogenic CO emissions may alter the concentration of various trace gases such as methane, hydrogen, and freons 21 and 22.

We have calculated the effects of increased CO emissions on tropospheric CH_4 levels. Alterations of the CH_4 concentration by increasing CO fluxes are of particular interest since (1) CH_4 is itself a major source of carbon monoxide and changes in its concentration will comprise a feedback into the perturbation chemistry, (2) CH_4 has an infrared absorption band at 7.7 μm which contributes to the heating of the lower atmosphere and changes in its concentration may thus have climatic significance, [*Wang et al.*, 1976] and (3) changes in the tropospheric CH_4 abundance will be reflected in corresponding stratospheric changes with possible impact on stratospheric photochemistry.

An immediate consequence of increased CO abundance is to shift the odd hydrogen balance from OH to HO_2 via the reactions (reaction numbers correspond to those given in Table 1):

(R20) $\quad CO + OH \rightarrow CO_2 + H$

(R28) $\quad H + O_2 \xrightarrow{M} HO_2$

Since the reaction of methane with OH,

(R17) $\quad CH_4 + OH \rightarrow CH_3 + H_2O$

is the only known tropospheric sink for CH_4, such a shift in the odd hydrogen balance can be expected to lead to an increase in the methane concentration [*Chameides et al.*, 1977; *Sze*, 1977]. However, we have found that the extent and even the sign of the methane change is dependent on the assumed NO_x ($NO_x = NO + NO_2$) background.

It is difficult to infer mean global NO or NO_2 concentrations from reported measurements. Observations till 1974 have been summarized by *Ackerman* [1975]. Tropospheric nitric oxide measurements were not available at that time and the reported nitrogen dioxide measurements were found to range from less than 0.1 ppb to several ppb. More recently, *Noxon* [1975] has measured NO_2 absorption spectra in Colorado. This yielded an upper limit on NO_2 mixing ratio of 0.1 ppb. Using continuous chemiluminescence analysis, *Cox* [1977] found $NO \leq 0.2$ ppb and $NO_2 = 0.4$ ppb in background maritime air in S.W. Ireland. *McElroy and Kerr* [1977] observed NO_2 absorption in limb scan experiments in Saskatchewan and obtained mixing ratios close to 1 ppb throughout the height of troposphere. Thus measurements of unpolluted tropospheric air indicate values of NO_2 concentration that differ by factors of greater than 10.

The importance of NO_x in the CO-CH_4 perturbation arises because the fate of the HO_2 molecule produced in reaction (R28) depends on the concentration of NO. The major sinks for the HO_2 molecule are given by the following reactions:

(R29) $\quad HO_2 + HO_2 \rightarrow H_2O_2 + O_2$

(R30a) $\quad HO_2 + NO \rightarrow NO_2 + OH$

(R33) $\quad HO_2 + O_3 \rightarrow OH + 2O_2$

Let us now consider two limiting cases. First, we assume that concentration of NO is vanishingly small so that the loss of HO_2 through (R30a) is nearly zero and HO_2 removal is dominated by (R29) and (R33). Thus the OH molecule lost in (R20) is either recovered in (R33) or is converted to H_2O_2 in (R29). This results in a net decrease in OH concentration and an increase in H_2O_2 concentration. This is the case reported by *Chameides et al.* [1977] and *Sze* [1977].

Next we consider the limiting case in which NO is so abundant that (R30a) dominates HO_2 removal in comparison with (R29) and (R33). The concentration of NO for which this would occur depends on the rates of reactions (R29), (R30a), and (R33). With the rates given in Table 1 we find that when $[NO_2] = 0.5$ ppb and $[O_3] = 45$ ppb, our model predicts [NO] = 0.14 ppb and $[HO_2] = 1.0 \times 10^{-3}$ ppb. In this case HO_2 loss

Paper number 8C0804
0148-0227/79/008C-0840$01.00

TABLE 1a. Reactions and Rates Used in the Present Model for Photolysis Reactions

Reaction No.	Reaction	Rate, s^{-1}	Reference	Wavelength Range, Å
(R1a)	$O_2 + h\nu \rightarrow O_2 + O$	1.2×10^{-6}	*Leighton* [1961] *Leighton* [1961] (cross section)	2800–7300
(R1b)	$O_2 + h\nu \rightarrow O_2 + O(^1D)$	1.5×10^{-6}	*Moortgat and Warneck* [1975] (quantum yields)	2800–3300
(R2)	$NO_2 + h\nu \rightarrow NO + O$	2.4×10^{-3}	*Johnston and Graham* [1974]	2800–4200
(R3a)	$NO_3 + h\nu \rightarrow NO + O_2$	1.2×10^{-2}*		
(R3b)	$NO_3 + h\nu \rightarrow NO_2 + O$	3.6×10^{-2}*		
(R4)	$N_2O_5 + h\nu \rightarrow NO_2 + NO_3$	3.4×10^{-5}	*Johnston and Graham* [1974]	2800–3800
(R5)	$HNO_3 + h\nu \rightarrow OH + NO$	$0.3J_2$	*Cox* [1975]	3000–4000
(R6)	$HNO_3 + h\nu \rightarrow OH + NO_2$	8.5×10^{-6}	*Johnston and Graham* [1974]	2800–3250
(R7)	$H_2O_2 + h\nu \rightarrow 2OH$	2.3×10^{-6}	*Leighton* [1961]	2800–3700
(R8a)	$H_2CO + h\nu \rightarrow HCO + H$	3.4×10^{-5}	*Calvert et al.* [1972]	2800–3600
(R8b)	$H_2CO + h\nu \rightarrow CO + H_2$	1.5×10^{-5}	*Calvert et al* [1972]	2800–3600
(R9)	$CH_3OOH + h\nu \rightarrow CH_3O + OH$	2.3×10^{-6}	Assume $k = k_7$ following *Chameides and Walker* [1976]	

*Calculated for zenith angle equal to zero by H. S. Johnston and R. Graham (personal communication, 1976).

rates through (R29), (R30a), and (R33) are $2k_{29}[HO_2] = 0.005$/s, $k_{30a}[NO] = 1.1$/s, and $k_{33}[O_3] = 0.03$/s, respectively. Therefore, for such NO_x concentrations, almost all OH molecules destroyed in (R20) are recovered in (R30a). Actually, an excess of OH results because the overwhelming majority of NO_2 molecules thus produced is photodissociated to NO and O, which results in ozone formation:

(R2) $$NO_2 + h\nu \rightarrow NO + O$$

(R10) $$O + O_2 + M \rightarrow O_3 + M$$

The O_3 molecule produced in (R10) can be removed in a number of ways which include interactions with NO_x, HO_x, and solar radiation. However, the sink mechanism which is significant in controlling OH abundance is (R1b), in which a small fraction of O_3 molecules is photodissociated, yielding an excited atom $O(^1D)$:

(R1b) $$O_3 + h\nu \rightarrow O_2 + O(^1D)$$

Most of $O(^1D)$ atoms are quenched to the ground state resulting in O_3 reformation by (R10). However, a few percent of $O(^1D)$ atoms interact with water vapor to yield two OH molecules:

(R12) $$O(^1D) + H_2O \rightarrow OH + OH$$

Thus for large NO_x concentrations the injection of CO into the atmosphere results in a net production of OH molecules, i.e., the perturbation in the OH concentration (and hence in the concentration of CH_4) changes sign as we go from low NO_x concentrations to high NO_x concentrations. Calculations illustrating this behavior are presented in the next section.

2. Calculations

We calculate the effects of variations in the CO flux into the atmosphere with a zero-dimensional equilibrium model of the troposphere [*Stewart et al.*, 1977]. The model computes the concentrations of trace species, for mean tropospheric conditions (i.e., density of air = 1.7×10^{19} molecules/cm^3, temperature = 273°K) by equating the rates of production and destruction in an iterative procedure which continues until fractional changes in all the computed concentrations are less than a prescribed 'error criterion.' Since we expect the perturbation due to CO injections to go through a qualitative change, the error criterion was taken as 10^{-6} to assure numerical accuracy in the critical region where the response of the model atmosphere changes sign. On the average a few thousand iterations per model computation were required to meet this criterion. The reaction rates are given in Tables 1a and 1b and are taken from the NASA publication 1010 [*Hudson*, 1977] unless otherwise stated.

The concentrations of CH_4, H_2, CO, NO_y, and O_3 are prescribed in the model at the estimated global averages (Table 2). The photochemical rate of production P_i of each of these species is augmented by a flux-divergence term F_i which is defined as flux of i into the troposphere minus flux of i out of the troposphere. F_i is adjusted so that $P_i + F_i = L_iC_i$, where L_i and C_i are the rate of loss and the concentration of the species i. The F_i calculated in this manner depend on the prescribed concentrations and the reaction rate coefficients. They also depend on the assumed NO_x abundance (only one of NO and NO_2 concentrations need be prescribed; the other is determined by the chemistry). The F_i obtained for $NO_y = 0.02$ ppb are shown in Table 1. The flux divergences from the model have been multiplied by the volume of the troposphere to obtain the units in megatons per year. For ozone a sink at the ground is included with a deposition velocity of 0.08 cm/s which means that the ground accounts for the loss of 521 Mt of ozone per year. Thus $F_{O_3} = 383$ Mt/yr means that transport from the stratosphere should be 904 Mt/yr to maintain the concentration of O_3 at 4.5 pphm.

The predicted variation in atmospheric methane as a function of CO flux into the atmosphere is shown in Figure 1. For each curve shown in this figure an unperturbed atmosphere was first calculated for a particular NO_y concentration and the other prescribed concentrations as given in Table 1. The perturbation study was then carried out by repeating this model calculation with the CO flux changed. Further details of the perturbation calculation are given in Table 3. In this table, concentrations with subscript zero are for the unperturbed atmosphere, and concentrations without this subscript are for the perturbed calculation in which CO flux into the atmosphere was increased by a factor of 2.

An examination of Figure 1 and Table 3 shows that the response of the model atmosphere to a CO flux perturbation changes sign near $[NO_y]_0 = 0.23$ ppb. For $[NO_y]_0 < 0.23$ ppb $[CH_4]$ increases with an increase in CO flux, while for $[NO_y]_0 \geq 0.23$ ppb $[CH_4]$ decreases with an increase in CO flux. This is because an increase in CO flux decreases [OH] for small $[NO_y]_0$ but increases [OH] for large $[NO_y]_0$. Molecular hydrogen is seen to change in a manner similar to that of methane.

TABLE 1b. Reactions and Rates Used in the Present Model for Nonphotolytic Reactions

Reaction No.	Reaction	Rate*	Reference
(R10)	$O + O_2 \xrightarrow{M} O_3$	$1.1 \times 10^{-34} e^{510/T}$	*Hampson and Garvin* [1975]
(R11a)	$O(^1D) \xrightarrow{N_2} O(^3P)$	$2.0 \times 10^{-11} e^{107/T}$	
(R11b)	$O(^1D) \xrightarrow{O_2} O(^3P)$	$2.9 \times 10^{-11} e^{67/T}$	
(R12)	$O(^1D) + H_2O \rightarrow 2OH$	2.3×10^{-10}	
(R13a)	$O(^1D) + CH_4 \rightarrow CH_3 + OH$	1.3×10^{-10}	
(R13b)	$O(^1D) + CH_4 \rightarrow H_2CO + H_2$	1.4×10^{-11}	
(R14)	$O(^1D) + H_2 \rightarrow H + OH$	9.9×10^{-11}	
(R15a)	$OH + OH \rightarrow O + H_2O$	$1.0 \times 10^{-11} e^{-550/T}$	
(R15b)	$OH + OH \xrightarrow{M} H_2O_2$	$1.25 \times 10^{-30} e^{500/T}$	
(R16)	$OH + HO_2 \rightarrow H_2O + O_2$	3.0×10^{-11}	
(R17)	$OH + CH_4 \rightarrow CH_3 + H_2O$	$2.4 \times 10^{-12} e^{-1710/T}$	
(R18)	$OH + CH_3OOH \rightarrow CH_3O_2 + H_2O$	$= k_{60}$	*Chameides and Walker* [1976]
(R19)	$OH + H_2CO \rightarrow HCO + H_2O$	$3.0 \times 10^{-11} e^{-500/T}$	
(R20)	$OH + CO \rightarrow H + CO_2$	1.4×10^{-13}	
(R21)	$OH + H_2 \rightarrow H + H_2O$	$1.8 \times 10^{-11} e^{-2000/T}$	*Smith and Zellner* [1974]§
(R22)	$OH + H_2O_2 \rightarrow HO_2 + H_2O$	$1.0 \times 10^{-11} e^{-550/T}$	
(R23)	$OH + NO \xrightarrow{M} HNO_2$	5.6×10^{-31}¶	*Cox* [1975]
(R24)	$OH + NO_2 \xrightarrow{M} HNO_3$	1.1×10^{-11}§	From t dependent expression in *Hudson* [1977]
(R25)	$OH + HNO_3 \rightarrow NO_3 + H_2O$	8.0×10^{-14}	
(R26)	$OH + HNO_2 \rightarrow NO_2 + H_2O$	2.1×10^{-12}	*Cox* [1975]
(R27)	$OH + O_3 \rightarrow HO_2 + O_2$	$1.5 \times 10^{-12} e^{-1000/T}$	
(R28)	$H + O_2 \xrightarrow{M} HO_2$	$2.1 \times 10^{-30} e^{300/T}$	
(R29)	$HO_2 + HO_2 \rightarrow H_2O_2 + O_2$	2.5×10^{-12}	
(R30a)	$HO_2 + NO \rightarrow OH + NO_2$	8.0×10^{-12}	
(R30b)	$HO_2 + NO \xrightarrow{M} HNO_3$	1.4×10^{-31}¶	*Cox* [1975]
(R31)	$HO_2 + NO_2 \rightarrow HNO_2 + O_2$	3.0×10^{-13}	*Cox* [1975]
(R32)	$HO_2 + NO_2 \rightarrow HNO_3 + O_2$	1.7×10^{-12}	*Demerjian et al* [1974]
(R33)	$HO_2 + O_3 \rightarrow OH + 2O_2$	$7.3 \times 10^{-14} e^{-1270/T}$	
(R34)	$HO_2 + H_2CO \rightarrow H_2O_2 + HCO$	$1.7 \times 10^{-12} e^{-800/T}$	
(R35)	$HO_2 + CH_3O_2 \rightarrow CH_3OOH + O_2$	6.7×10^{-14}	*Demerjian et al* [1974]
(R36a)	$CH_3 + O_2 \xrightarrow{M} CH_3O_2$	4.0×10^{-13}¶	*Demerjian et al* [1974]
(R36b)	$CH_3 + O_2 \rightarrow H_2CO + OH$	$2.9 \times 10^{-12} e^{-500/T}$	*Washida and Bayes* [1976]||
(R37)	$2CH_3O_2 \rightarrow 2CH_3O + O_2$	1.7×10^{-13}	*Parkes et al* [1973]||, *Demerjian et al* [1974]
(R38)	$CH_3O + O_2 \rightarrow H_2CO + HO_2$	$1.6 \times 10^{-13} e^{-2000/T}$	*Heicklen* [1973]||
(R39)	$HCO + O_2 \rightarrow CO + HO_2$	6.0×10^{-12}	
(R40)	$CH_3O_2 + NO \rightarrow CH_3O + NO_2$	$3.3 \times 10^{-12} e^{-300/T}$	*Demerjian et al* [1974]
(R41)	$CH_3O_2 + NO_2 \rightarrow HNO_2 + H_2CO$	$0.05 k_{60}$	*Cox et al* [1976]
(R42)	$NO_2 + O_3 \rightarrow NO_3 + O_2$	$1.2 \times 10^{-13} e^{-2450/T}$	
(R43)	$NO + O_3 \rightarrow NO_2 + O_2$	$2.1 \times 10^{-12} e^{-1450/T}$	
(R44a)	$NO_2 + NO_3 \xrightarrow{M} N_2O_5$	3.8×10^{-12}**	*Demerjian et al.* [1974]
(R44b)	$NO_2 + NO_3 \rightarrow NO_2 + NO + O_2$	$2.3 \times 10^{-13} e^{-1000/T}$	
(R45)	$N_2O_5 \xrightarrow{M} NO_2 + NO_3$	$5.7 \times 10^{14} e^{-10,000/T}$‡	*Baulch et al.* [1973]||
(R46)	$NO_3 + NO \rightarrow 2NO_2$	8.7×10^{-12}	
(R47a)	$HNO_3 \rightarrow$ rainout	1.15×10^{-6}	10-days rainout time
(R47b)	$H_2O_2 \rightarrow$ rainout	1.15×10^{-6}	
(R47c)	$H_2CO \rightarrow$ rainout	1.15×10^{-6}	
(R47d)	$CH_3OOH \rightarrow$ rainout	1.15×10^{-6}	
(R48)	$N_2O_5 + H_2O \rightarrow 2HNO_3$	$1.0 \times 10^{-20} e^{-10,000/T}$	
(R49)	$HO_2 + CO \rightarrow OH + CO_2$	1.0×10^{-20}	

*Rates are from Hudson [1977] unless otherwise indicated. Unimolecular, bimolecular, and termolecular rate constants are given in units of s^{-1}, $cm^3\ s^{-1}$, and $cm^6\ s^{-1}$. See also footnotes below.

¶Measurement at 1-atm pressure.

§Second-order high-pressure limit.

||Quoted in *Hampson and Garvin* [1975].

**Calculated from reverse rate.

‡First-order high-pressure limit.

TABLE 2. Flux Divergences F, Computed in the Model to Yield the Prescribed Species Concentrations

Species	Concentration, ppm	Flux Divergence, MT/yr	Other Estimates of Flux, MT/yr	Reference
CH_4	1.4	278	546–1060	[*Enhalt*, 1974]
H_2	0.5	−10.7	48	[*Schmidt*, 1974]
CO	0.125	512	359,640	[*Jaffe*, 1975], [*Seiler*, 1974]
NO_y	2.0×10^{-8}	122		
O_3	4.5×10^{-8}	383		Import from stratosphere estimated at 900 MT/yr [*Danielsen and Mohnen*, 1977]

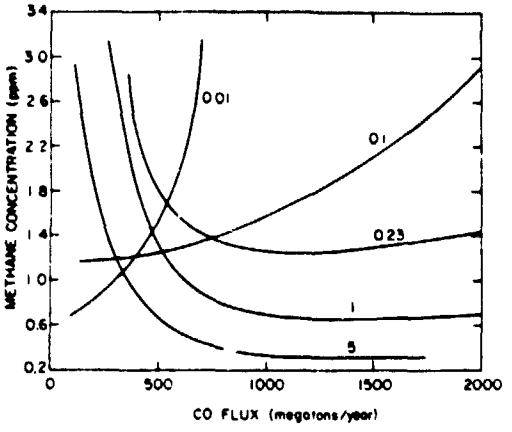


Fig. 1 Predicted equilibrium values of methane concentration as a function of carbon monoxide input into the atmosphere. The numbers on the curves indicate the concentration of O_3 (in ppb) in the unperturbed atmosphere containing 1.4 ppm of methane.

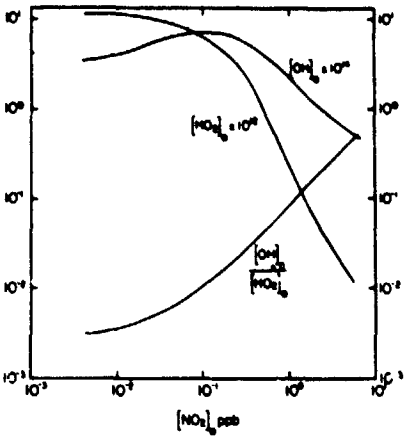


Fig. 2. Variation of OH and HO_2 concentrations as functions of NO_2 concentrations in the unperturbed atmosphere. Notice that as $[NO_2]_0$ increases, $[OH]_0$ increases at first and then decreases while $[HO_2]$ decreases monotonically. For large $[NO_2]_0$ the rate of $[HO_2]_0$ decrease is faster than that of $[OH]_0$ resulting in a steady increase in the ratio $[OH]_0/[HO_2]_0$.

3. Discussion

The critical nature of the assumed background NO_2 concentration is manifested also in the hydroxyl concentration $[OH]_0$ calculated in the unperturbed model. As can be seen in Table 3 and Figure 2, $[OH]_0$ is not a linear function of $[NO_2]_0$ but passes through a maximum. Thus calculated CO, CH_4, and H_2 lifetimes pass through minima as functions of $[NO_2]_0$.

The mean hydroxyl radical concentration and its variations in the troposphere have been extensively discussed in recent literature [*Crutzen and Fishman*, 1977; *Liu*, 1977, and references therein]. This is because the abundance of OH radical is important for the removal of both naturally occurring trace gases as well as those released by anthropogenic activity. The hydroxyl radical is probably the most effective cleansing agent in the atmosphere, and understanding its atmospheric interactions is important. *Liu* [1977] discussed the possibility of increases in atmospheric OH concentrations as a result of increased NO emissions. Because of short NO and NO_2 atmospheric lifetimes it is not clear if increased anthropogenic NO emissions would result in significantly higher NO_x concentrations in the troposphere as a whole. However, if there is an increase in the ambient NO_x concentrations, Figure 2 shows that Liu's result would obtain at low levels of $[NO_x]$. Further increases in NO_x concentrations would eventually lead to a reversal in the hydroxyl concentration response.

Our model calculations suggest the following reasons for the variation of $[OH]_0$ with respect to $[NO_2]_0$ seen in Figure 2. There are two major removal mechanisms for odd hydrogen radicals HO_x ($HO_x = OH + HO_2 + 2H_2O_2$). One is the rainout of H_2O_2:

(R47*b*) H_2O_2 rainout; $k_{14} = 1.15 \times 10^{-6}\ s^{-1}$

and the other is conversion to nitric acid and its rainout:

(R24) $NO_2 + OH \xrightarrow{M} HNO_3$; $k_{24} = 1.14 \times 10^{-11}$ cm^3 mol^{-1} s^{-1}

(R30*b*) $NO + HO_2 \xrightarrow{M} HNO_3$; $k_{30b} = 1.4 \times 10^{-13}$ cm^3 mol^{-1} s^{-1}

(R47*a*) HNO_3 rainout; $k_{14a} = 1.15 \times 10^{-6}\ s^{-1}$

NO also converts HO_2 to OH:

(R30*a*) $NO + HO_2 \rightarrow NO_2 + OH$;

$k_{30a} = 8.0 \times 10^{-12}$ cm^3 mol^{-1} s^{-1}

When $[NO_2]_0$ is very small, the share of (R24), (R30*b*), and (R47*a*) in HO_x removal is also very small (see Table 4 for rates of removal). Hence a small increase in $[NO_2]_0$ decreases the HO_x density (and therefore HO_2 density) only slightly. Since increasing $[NO_2]_0$ increases $[NO]_0$ by the same factor, the conversion of HO_2 to OH by (R30*a*) becomes more effective.

TABLE 3 Results of the Perturbation Calculations

$[H_2]$, ppm	$[NO_2]_0$, ppb	$[CH_4]$, ppm	$[CO]$, ppm	$[OH]_0$, 10^{-5} ppb	$[OH]$, 10^{-5} ppb	$[HO_2]_0$, 10^{-5} ppb	$[HO_2]$, 10^{-5} ppb	$[O_3]$, pphm
65.1	0.005	105.0	15.6	3.6	0.047	11.3	13.7	4.27
5.3	0.01	9.2	1.4	3.9	0.59	11.2	13.5	4.53
0.96	0.1	2.14	0.32	6.6	4.3	7.1	10.2	5.78
0.56	0.2	1.50	0.22	6.6	6.2	3.9	7.5	6.15
0.47	0.23	1.36	0.20	6.4	6.5	3.3	6.9	6.19
0.15	0.5	0.72	0.11	4.1	7.9	1.0	3.5	6.23
0.10	1.0	0.56	0.08	2.3	5.7	0.27	1.2	6.15
0.29	3.0	0.91	0.14	0.88	1.4	0.035	0.082	5.36
0.36	5.0	1.1	0.16	0.58	0.76	0.014	0.026	5.10

Subscript zero indicates the concentration is for unperturbed atmosphere in which $[CH_4]_0 = 1.4$ ppm, $[CO]_0 = 0.125$ ppm, $[O_3]_0 = 4.5$ pphm, and $[H_2]_0 = 0.5$ ppm. Concentrations without subscripts are obtained when the CO flux into the atmosphere is doubled.

TABLE 4. Sensitivity of HO_x to NO_x in the Unperturbed Model

	Small $[NO_x]_0$		Large $[NO_x]_0$	
	$[NO_x]_0$ = 0.005 ppb	$[NO_x]_0$ = 0.01 ppb	$[NO_x]_0$ = 0.5 ppb	$[NO_x]_0$ = 1 ppb
HO_x Concentration, ppm	2.8×10^{-8}	2.7×10^{-8}	2.2×10^{-8}	2.0×10^{-8}
HO_x Removal Rate Due to (R14), $ppm^{-1}\ min^{-1}$	6.9×10^{-6}	6.9×10^{-6}	6.6×10^{-6}	5.9×10^{-6}
HO_x Removal Rate Due to (R15)–(R17), $ppm^{-1}\ min^{-1}$	1.5×10^{-6}	3.3×10^{-6}	1.3×10^{-5}	1.6×10^{-5}
OH Concentration, ppm	3.6×10^{-6}	3.9×10^{-6}	4.1×10^{-6}	2.3×10^{-6}

Note that for small NO_x concentrations, increasing $[NO_x]_0$ by a factor of 2 results in a very small decrease in HO_x concentration, and therefore [OH] increases because of (R18). The reverse is true for large NO_x concentrations.

and a net increment in the OH concentration results. This behavior is reversed in the regime of large $[NO_x]_0$. Removal through (R24), (R30*b*), and (R47) is the dominant HO_x sink and a small increase in $[NO_x]_0$ results in a significant decrease in HO_x densities. This decrease is so large that a net reduction in OH concentration results in spite of (R30*a*). Note also, as shown in Figure 2, that HO_x concentration is a decreasing function of $[NO_x]_0$. For large values of NO_x concentration the rate of decrease of HO_x exceeds that of OH so that the ratio $[OH]_0/[HO_x]_0$ remains an increasing function of $[NO_x]_0$.

Although we have not carried out a time dependent calculation, these results on OH concentrations show that the time scales of the CO-CH_4 perturbation also depend on the NO_x background concentrations.

4. Conclusion

Our calculations indicate that changes in the tropospheric concentrations of CH_4 (and other species such as CO, OH, H_2, O_3, etc.) due to variation in the CO input into the atmosphere depend, both in magnitude and sign, on the concentrations of NO and NO_2. Knowledge of the tropospheric distributions of nitrogen oxides is therefore critical for studying the effects of changes in the CO flux. Measurements of NO_x at different locations have yielded large differences in its concentrations [*Ackerman*, 1975; *Noxon*, 1975; *McElroy and Kerr*, 1977]. Unless there are systematic errors in some of these measurements, we can conclude that NO_x distribution in the troposphere is highly nonhomogeneous. The results presented in this paper then indicate that the effects of CO fluxes into the atmosphere, being controlled by ambient NO_x, proceed differently at different locations. Because this modulation by NO_x makes the impact of CO fluxes on OH and other species concentrations nonlinear, an evaluation of the global impact of increasing CO input into the atmosphere should take account of the nonuniformity in NO_x distributions.

Furthermore, it has been suggested that because of anthropogenic CO fluxes, the present concentrations of methane and molecular hydrogen may be substantially larger than those prevailing in the pre-industrial era [*Crutzen and Fishman*, 1977]. The results presented in this paper show that, because of the NO_x relationship with the CO-CH_4 coupling, this conclusion is not necessarily correct. In order to deduce pre-industrial concentrations of CH_4 and H_2 from photochemical calculations we would have to know the NO_x concentration distribution at that time. This would require estimates of NO_x production rates in the biosphere and in lightning discharges at the time in question.

Acknowledgment. This research was supported in part by NASA grant NSG 5102 to State University of New York at Stony Brook.

References

Ackerman, M., NO, NO_2 and HNO_3 below 35 km in the atmosphere, *J. Atmos. Sci.*, *32*, 1649–1657, 1975.

Baulch, D. L., D. D. Drysdale, and D. G. Horne, Rate constants for reactions of the H_2-N_2-O_2 system, Chemical Kinetics Data Survey V, *NBSIR 73-206*, pp. 49–115, Nat. Bur. of Stand., Washington, D. C., 1973.

Calvert, J. G., J. A. Kerr, K. L. Demerjian, and R. D. McQuigg, Photolysis of formaldehyde as hydrogen atom source in the lower atmosphere, *Science*, *175*, 751–752, 1972.

Chameides, W. L., and J. C. G. Walker, A time-dependent photochemical model for ozone near the ground, *J. Geophys. Res.*, *81*, 413–420, 1976.

Chameides, W. L., S. C. Liu, and R. J. Cicerone, Possible variation in atmospheric methane, *J. Geophys. Re.*, *82*, 1795–1798, 1977.

Cox, R. A., The photolysis of gaseous nitrous acid—A technique for obtaining kinetic data on atmospheric photo-oxidation reactions, in *Proceedings of the Symposium on Chemical Kinetics Data for the Lower and Upper Atmosphere*, edited by S. W. Benson, pp. 379–398, Interscience, New York, 1975.

Cox, R. A., Some measurements of ground level NO, NO_2 and O_3 concentrations at an unpolluted maritime site, *Tellus*, *29*, 356–362, 1977.

Cox, R. A., R. G. Derwent, P. M. Holt, and A. Kerr, Photo-oxidation of methane in the presence of NO and NO_2, *J. Chem. Soc. Faraday Trans. 1*, *72*, 2044–2060, 1976.

Crutzen, P. J., and J. Fishman, Average concentrations of OH in the troposphere and the budgets of CH_4, CO, H_2, and CH_3CCl_3, *Geophys. Res. Lett.*, *4*, 321–324, 1977.

Danielsen, E. F., and V. A. Mohnen, Project dustorm report: Ozone transport, in situ measurements, and meteorological analyses of tropopause folding, *J. Geophys. Res.*, *82*, 5867–5877, 1977.

Demerjian, K. L., J. A. Kerr, and J. G. Calvert, The mechanism of photochemical smog formation, *Advan. Environ. Sci. Technol.*, *4*, 1–262, 1974.

Enhalt, D. H., The atmospheric cycle of methane, *Tellus*, *26*, 58–70, 1974.

Hampson, R. F., and D. Garvin, Chemical kinetic and photochemical data for modelling atmospheric chemistry, *Tech. Note 666*, 113 pp., Nat. Bur. of Stand., Washington, D. C., 1975.

Heicklen, J., Photochemical and rate data for methylnitrite, methoxy and methylperoxy, Chemical Kinetics Data Survey V, *NBSIR 73-206*, pp. 43–48, Nat. Bur. of Stand., Washington, D. C., 1973.

Hudson, R. D. (Ed.), Chlorofluoromethanes and the stratosphere, *NASA Ref. Publ. 1010*, Nat. Aeron. and Space Admin., Washington, D. C., 1977.

Jaffe, L. S., The global balance of carbon monoxide, in *The changing Global Environment* edited by S. F. Singer, pp. 83–110, D. Reidel, Hingham, Mass., 1975

Johnston, H. S., and R. Graham, Photochemistry of NO_x and HNO_x compounds, *Can. J. Chem.*, *52*, 1415–1423, 1974

Leighton, P. A., *Photochemistry of Air Pollution*, Academic, New York, 1961

Liu, S. C., Possible effects on tropospheric O_3 and OH due to NO emissions, *Geophys. Res. Lett.*, *4*, 325–328, 1977

McElroy, C. T., and J. B. Kerr, A limb scan measurement of atmospheric nitrogen dioxide, paper presented at the Fall Meeting, AGU, San Francisco, Calif., 1977

Moortgat, G. K., and P. Warneck, Relative $O(^1D)$ quantum yields in the near UV photolysis of ozone at 298 K, *Z. Natur.*, *A30*, 835–844, 1975

Noxon, J. F., Nitrogen dioxide in the stratosphere and troposphere measured by the ground based spectroscopy, *Science*, *189*, 547–549, 1975

Parkes, D. A., D. M. Paul, C. P. Quinn, and R. C. Robson, The ultraviolet absorption by alkyl peroxy radicals and their mutual reactions, *Chem. Phys. Lett.*, *23*, 425–429, 1973

Schmidt, U., Molecular hydrogen in the atmosphere, *Tellus*, *26*, 78–90, 1974

Seiler, W., The cycle of atmospheric CO, *Tellus*, *26*, 116–135, 1974

Smith, I. W. M., and R. Zellner, Rate Measurements of reactions of OH by resonance absorption, 3, Reactions of OH with H_2, D_2, hydrogen and deuterium halides, *J. Chem. Soc. Faraday Trans 2*, *70*, 1045–1056, 1974

Stewart, R. W., S. Hameed, and J. Pinto, Photochemistry of tropospheric ozone, *J. Geophys. Res.*, *82*, 3134–3139, 1977.

Sze, N. D., Anthropogenic CO emission Implications for atmospheric CO-OH-CH_4 cycle, *Science*, *195*, 673–675, 1977

Wang, W. C., Y. L. Yung, A. A. Lacis, T. Mo, and J. E. Hansen, Greenhouse effects due to man-made perturbations of trace gases, *Science*, *194*, 685–690, 1976.

Washida, N., and K. D. Bayes, The reactions of methyl radicals with atomic and molecular oxygen, *Int. J. Chem. Kinet.*, *8*, 777–794, 1976

(Received August 29, 1977,
revised June 29, 1978,
accepted July 14, 1978.)

VOL. 6, NO. 11 GEOPHYSICAL RESEARCH LETTERS NOVEMBER 1979

LATITUDINAL DISTRIBUTION OF THE SOURCES OF CARBON MONOXIDE IN THE TROPOSPHERE

Sultan Hameed

Laboratory for Planetary Atmospheres Research, Department of Mechanical Engineering
State University of New York at Stony Brook, Stony Brook, Long Island, N. Y. 11794

Richard W. Stewart

Laboratory for Atmospheric Sciences, Goddard Space Flight Center, NASA, Greenbelt, Maryland 20770

Abstract. We have constructed a vertically and zonally averaged model of the troposphere which calculates photochemical interactions and diffusive North-South transport of trace species. The model can be used to calculate the latitudinal distribution of the source function of a species if its concentration distribution is known. We have applied this procedure to carbon monoxide and find large sources outside the industrialized belt in the Northern Hemisphere.

Introduction

Carbon monoxide is produced in the troposphere by large natural and anthropogenic sources. The source due to fossil fuel burning has been estimated to be 640 MT/yr. by Seiler (1974, 1975). The existence of large natural sources of CO was deduced by Weinstock and Niki (1972) by a budget analysis of radiocarbon. Since then, oxidation of methane by hydroxyl radicals has been widely accepted as such a source. Photochemical model calculations give magnitudes of CO production from methane ranging from 300 to 800 MT/yr. depending upon the concentration of OH in the particular model (see e.g. Chameides and Cicerone, 1978; Crutzen and Fishman, 1977; Hameed et al. 1979). Oxidation of non-methane hydrocarbons in the atmosphere has also been suggested as a significant source of CO. Chameides and Cicerone (1978) calculate this source to be 220 MT/y.. Zimmerman et al. (1978) estimate this source to be 420 to 1330 MT/yr. and attribute it to oxidation of isoprene and terpenes. In this paper, we estimate the total CO source and its latitudinal distribution required to maintain the observed CO concentration distribution by a model calculation of photochemical interactions and transport in the troposphere. Our results show that large natural sources of CO exist, especially in tropical latitudes. These sources are in addition to methane oxidation which we explicitly calculate.

Model Calculations

The zonally and vertically averaged annual mean mixing ratios, C_i, for the species i, are obtained from numerical solution of the continuity equations (Czeplak and Junge, 1974):

$$\frac{1}{R^2}\frac{d}{d\mu}\left[K(\mu)\,(1-\mu^2)\,\frac{d}{d\mu}\,C_i\right] + \frac{1}{R^2}\,\frac{K(\mu)}{h(\mu)}\,(1-\mu^2)\cdot\frac{dh}{d\mu}\,\frac{dC_i}{d\mu} + P_i\,(C_1,\ldots C_i\ldots C_N) + Q_i\,(\mu) - C_i\cdot\left[L_i\,(C_1,\ldots C_i\ldots C_N) + S_i\,(\mu)\right] = 0,\quad i = 1, 2, \ldots, N \qquad (1)$$

where $\mu = \sin\theta$, θ is latitude, K is the north-south diffusion coefficient, h is the tropopause height and R is earth's radius. The photochemical production and destruction rates for species i are P_i and L_i respectively and depend on the concentrations of other species. Those sources and sinks which are not computed from photochemical interactions of species are denoted by Q_i and S_i, respectively. Examples of these are anthropogenic emissions and destruction by contact with earth's surface.

The set of coupled equations (1) is solved numerically by iteration with the boundary conditions that the flux of each species at the poles is zero. The region of integration is divided into 18 boxes centered at 85°S, 75°S - - - - 75°N, 85°N. If the source and sink functions $Q_i(\mu)$ and $S_i(\mu)$ are known, equations (1) can be solved for the steady state concentrations C_i. However, if the latitudinal distribution of the concentration of a species has been measured and its sinks can be estimated, the equation can be used to compute the source function:

$$Q_i(\mu) = C_i\left[L_i + S_i\right] - P_i - \frac{1}{R^2}\,\frac{d}{d\mu}\left[K(1-\mu^2)\,\frac{dC_i}{d\mu}\right] + \frac{1}{R^2}\,\frac{K}{h}\,(1-\mu^2)\,\frac{dh}{d\mu}\,\frac{dC_i}{d\mu} \qquad (2)$$

The CO mixing ratio as a function of latitude has been measured over the Atlantic and Pacific Oceans by Seiler (1975) to an accuracy of 1 ppb. The vertically and zonally averaged CO mixing ratios given by Seiler are used here to compute the CO source function using equation (2).

The calculations are performed for annually averaged conditions. The tropopause height in the model varies from 8 km. at the poles to 16 km. at the equator. The north-south diffusivity coefficient K is taken from curve 1b of Fig. 1 of Czeplak and Junge (1974). It is based on wind variance data from Newell et al. (1972). The diffusivity function has a maximum at 55° latitude (3.7×10^{10} cm^2 sec^{-1}) and its value at the equator is 1.9×10^{10} cm^2 sec^{-1}. K is taken as the same for the two hemispheres. The

Paper number 9L1285.
0094-8276/79/0091-1785,01.00

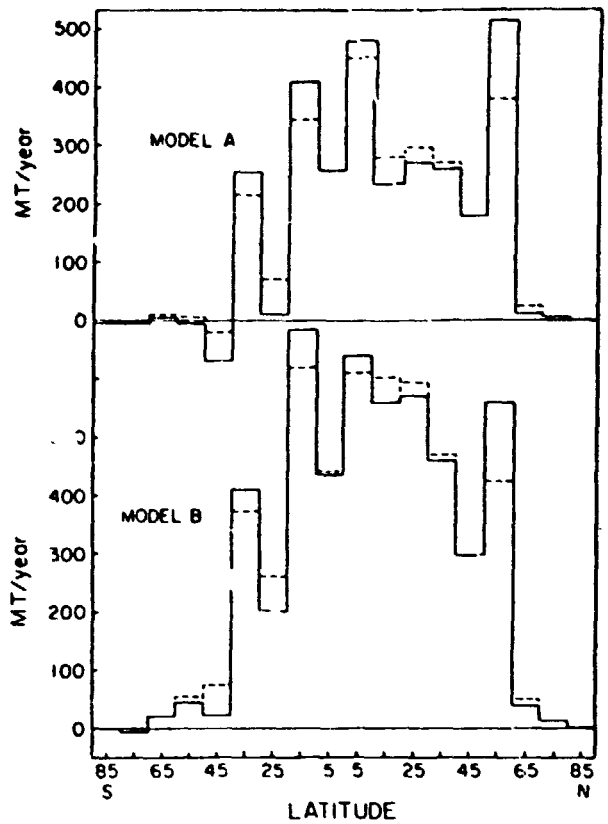


Fig. 1. Distribution of CO sources in addition to methane oxidation required to maintain the observed CO concentration distribution according to Model A (top) and Model B (bottom). OH concentrations in Model B are generally twice those in Model A.

set of photochemical reactions used in calculating production and loss terms P_i and L_i are those given in Tables 1a and 1b of Hameed et al. (1979). These reactions represent the interactions between NO_x and O_x species and those derived from oxidation of methane and water vapor. Other hydrocarbons are not considered. Values of reaction rates have been updated, according to the recommendations of Demore et al. (1979). The rates of photolysis reactions were calculated taking into account ozone absorption, Rayleigh scattering, aerosol scattering, ground albedo and mean cloud amounts as functions of latitude (Matloff and Stewart, 1979). Measured mixing ratios of methane (1.6 ppm in the Southern Hemisphere increasing to 1.7 ppm in the Northern Hemisphere), hydrogen (0.5 ppm everywhere) and ozone (Wilcox and Belmont, 1977) are also prescribed in the model and their sources computed in the same way as the CO source reported here. The emission of NO_x as a function of latitude is chosen so that the calculated mixing ratios of nitric acid are in agreement with the measurements of Huebert and Lazrus (1979) for the free troposphere.

The source strength Q depends upon the model values of K and the chemical production and destruction terms. The latter are proportional to OH concentrations which are sensitive to some poorly determined variables such as ambient NO_x, O_3 and water vapor concentrations and the photolysis rate of O_3 to O(1D). It has been suggested by Singh et al. (1979) that observed concentrations of methyl chloroform are consistent with OH concentrations which are significantly lower than model predictions. Due to this uncertainty we consider two versions of our model to estimate the CO source strength: Model A in which temperature, water vapor density and ozone are prescribed appropriate to the mean density level in each latitudinal zone, and Model B in which these quantities are given vertical mean values in each zone. In each case water vapor and temperature data for the Northern Hemisphere are taken from London (1957) and the corresponding numbers for the Southern Hemisphere are estimated from Van Loon et al. (1972). Representative OH concentrations (in 10^6 molecules/cm^3) for Model A are: 0.36 at 65°S, 1.3 at 35°S, 1.2 at 5°S, 1.5 at 5°N, 0.95 at 35°N and 0.12 at 65°N. For Model B: 0.72 at 65°S, 2.9 at 35°S, 2.1 at 5°S, 2.3 at 5°N, 1.8 at 35°N and 0.34 at 65°N. OH concentrations in Model B are larger mainly due to higher prescribed water vapor concentrations. The calculated source functions for these two models are shown in Figure 1. They show large emissions in the tropical regions and distinct peaks at 55°N, 5°N, 15°S and 35°S, instead of one peak usually expected for the CO source function (Seiler 1974). The dashed lines in the figure give the predicted source strengths when K is multiplied by 0.64. The diffusion coefficient is represented by the formula $K = \alpha^{-1}\sigma^2$ where σ is the variance of the meridional wind component and α is a constant. Czeplak and Junge (1974) used $\alpha = 2.4 \times 10^{-5}$ sec^{-1} while Newell et al. (1972) took $\alpha = 3.75 \times 10^{-5}$ sec^{-1}. Thus the dashed lines give the result for the value of α preferred by Newell et al.

The following processes determine the budget of CO in a latitude zone in our calculations.

1. Oxidation of CH_4 by OH which leads to reactions

   (1a) H_2CO + photon = CO + H_2, with latitude dependent photolysis rate,

   (1b) $HCO + O_2 = CO + HO_2$, $k_{1b} = 5 \times 10^{-12}$ cm^3 sec^{-1}.

2. Destruction due to reaction with OH: $CO + OH = CO_2 + H$; $k = 1.35 \times 10^{-13}(1 + M/Mo)$ cm^3 sec^{-1}, where M is the vertical mean density and Mo is density at surface level.

3. Transport to and from other latitudes due to meridional fluxes in equation (2).

4. Destruction by contact with earth's surface. The rate of loss is estimated from Seiler and Schmidt (1974).

5. The source function Q, to be determined by balancing the sum of the contributions of the other four processes.

The calculated values of these contributions to the CO budget according to Model A are shown in Table 1 in units of millions of tons per year. The second column gives the measured mean tropospheric mixing ratio (Seiler, 1975) and the last column gives the net source Q of CO required to maintain this mixing ratio. The last but one column lists the net gain of CO in the latitude zone due to transport. We note that the four latitudes where the computed source function has sharp peaks are net exporters of CO. This behavior is determined by the gradi-

TABLE 1. Contributions to the carbon monoxide budget for Model A in MT/yr.

Latitude	CO Concentration (ppm)	Production from Methane (Reac.(1a)	Reac.(1b)	Chemical Loss Reaction with OH	Loss at Earth's Surface	Gain by Transport	Net Source Q
85°S	0.055	.4	.2	-3	0	3	-1
75°S	0.056	1	.7	-10	0	19	-11
65°S	0.059	3	2	-22	-2	13	5
55°S	0.061	7	5	-50	-0.4	45	-7
45°S	0.064	16	14	-105	-2	150	-73
35°S	0.072	29	29	-196	-10	-106	254
25°S	0.076	31	29	-214	-19	161	12
15°S	0.096	34	30	-269	-23	-181	409
5°S	0.104	37	33	-315	-27	15	257
5°N	0.118	44	46	-460	-30	-82	482
15°N	0.116	40	38	-398	-34	121	233
25°N	0.125	33	49	-358	-50	56	270
35°N	0.134	26	23	-271	-66	27	261
45°N	0.143	13	9	-138	-81	19	178
55°N	0.154	6	4	-80	-82	-362	514
65°N	0.151	2	1	-21	-37	46	9
75°N	0.150	.7	.4	-9	0	3	5
85°N	0.149	.2	.1	-2	0	3	-1
	TOTAL	323	313	-2921	-463	-50	2797
TOTALS for Model B		542	839	-5800	-463	-50	4933

ents of the measured CO distribution. We also note that the total of this column is - 50 MT/yr. (it should be zero). This represents 3 per cent of the total CO mass transported and is a measure of numerical error in our finite difference scheme. When we multiply K by 0.64 (dashed lines in the figure) entries in this column are multiplied by 0.64 This results in a more evenly distributed source function Q.

We see that according to Model A the total source in addition to methane is 2797 MT/yr. Model B gives 4933 MT/yr. for the total source.

Conclusions

The anthropogenic sources of CO are estimated to be 640 MT/yr. due to fossil fuel burning and 60 MT/yr. due to non-fossil fuel burning (Seiler, 1974, 1975). These are located mostly in the industrialized regions between 25°N and 55°N. Our estimated sources at these latitudes are in excess of Seiler's estimates of anthropogenic sources and point to possible presence of natural CO sources in this region. The peaks at 55°N in the source functions in Fig. 1 are indicative of this because the anthropogenic source should have a maximum at lower latitudes. We find even larger sources in the equatorial regions and these, clearly, should be due to some natural processes. Recently, Zimmerman et al. (1978) have pointed out a hitherto unrecognized large source of carbon monoxide in the oxidation of heavy hydrocarbons (other than methane) which are released by vegetation. Based on data in the United States, they have estimated the inventory of isoprene and terpenes and calculate CO production from the oxidation of these species to be between 420 and 1330 MT/yr. Chameides and Cicerone (1978) have suggested significant sources of atmospheric CO due to the oxidation of other hydrocarbons. These calculations are in general agreement with our conclusion of the existence of large non-

methane natural sources of CO. Our model calculations indicate the latitudinal distribution of such sources.

Acknowledgment: This study was supported by NASA grant NSG 5102 to State University of New York at Stony Brook.

References

Chameides, W. L., and R. J. Cicerone, Effects of Nonmethane Hydrocarbons in the Atmosphere, J. Geophys. Res., 83, 947-952, 1978.

Crutzen, P. J., and J. Fishman, Average concentrations of OH in the troposphere, and the budgets of CH_4, CO, H_2 and CH_3CCL_3, Geophys. Res. Ltrs., 4, 321-324, 1977.

Czeplak, G., and C. Junge, Studies of interhemispheric exchange in the troposphere by a diffusion model, Advances in Geophysics, 18B, 57-72, 1974.

Demore, W. B., L. J. Stief, S. Kaufman, D. M. Golden, R. F. Hampson, M. J. Kurylo, J. J. Margitan, M. J. Molina, and R. T. Watson, Chemical Kinetics and Photochemical Data for use in Stratospheric Modeling, Jet Propulsion Laboratory Report 79-27, 1979.

Hameed, S., J. P. Pinto and R. W. Stewart, Sensitivity of the predicted CO-OH-CH_4 perturbation to tropospheric NO_x concentrations, J. Geophys. Res., 84, 763-768, 1979.

Huebert, B. J., and A. L. Lazrus, Tropospheric gas-phase and particulate nitrate measurements, J. Geophys. Res. (to appear), 1979.

London, J. A study of the atmospheric heat balance, Report Contract AF 19 (122-165) College of Engineering, New York University, 1957.

Matloff, G., and R. W. Stewart, Tropospheric UV flux calculation and photolysis rates for use with zonally and diurnally averaged models, Applied Optics, 1979, in press.

Newell, R. E., J. W. Kidson, D. G. Vincent, and J. G. Boer, The general circulation of the tropical atmosphere and interactions with extratropical latitudes, Vol. 1, MIT Press, Cambridge, Mass. 1972.

Seiler, W., The cycle of atmospheric CO, Tellus, 26, 116-135, 1974.

Seiler, W., Cycle of Carbon Monoxide in the Atmosphere, Proceedings of the International Conference on Environmental Sensing and Assessment, Las Vegas, Nevada, 1975.

Seiler, W., and U. Schmidt, New Aspects on CO and H_2 cycles in the Atmosphere, Proceedings of the International Conference on Structure, Composition and General Circulation of the Upper and Lower Atmosphere and possible anthropogenic perturbations, Melbourne, Vol. 1, 192-222, 1974.

Singh, H. B., L. J. Salas, H. Shigeishi, and E. Scribner, Atmospheric halocarbons, hydrocarbons and sulfur hexafluoride: global distributions, sources and sinks, Science, 203, 899-903, 1979.

Van Loon, J., J. J. Taljaard, T. Sasamori, J. London, D. V. Hoyt, K. Labitzke, C. W. Newton, Meteorology of the Southern Hemisphere, American Meteorological Society, Boston, Mass., 1972.

Weinstock, B., and H. Niki, Carbon monoxide balance in nature, Science, 176, 290-292, 1972.

Wilcox, R. W., and A. D. Belmont, Ozone concentration by latitude, altitude and month near 80°W, U. S. Department of Transportation Report No. FAA-AEQ-77-13, Washington, D. C. 20591, 1977.

Zimmerman, P. R., R. B. Chatfield, J. Fishman, P. J. Crutzen, and P. L. Hanst, Estimates on the production of CO and H_2 from the oxidations of hydrocarbon emissions from vegetation, Geophys. Res. Ltrs., 5, 679-681, 1978.

(Received May 17, 1979;
accepted September 4, 1979.)

D4

N81-21690

Paper 28

Reprinted from
(COSPAR) REMOTE SOUNDING OF THE ATMOSPHERE FROM SPACE
Edited by
H. J. BOLLE

PERGAMON PRESS OXFORD and NEW YORK 1979

COMPARISONS OF GLOBAL OZONE TRENDS INFERRED FROM THE BUV EXPERIMENT ON NIMBUS 4 AND THE GROUND - BASED NETWORK

D. F. Heath

NASA/Goddard Space Flight Center, Greenbelt, Maryland, U.S.A.

Preliminary comparisons between global ozone burdens derived from the Back-scattered Ultraviolet (BUV) experiment on Nimbus 4 and those inferred from an analysis of ground-based network data seem to indicate significant differences in the inter-annual variability of ozone. A study is made of the reasons for the differences. Some of the observed differences may by due to improper weighting of the groundbased network data, slowly changing planetary wave structure over the fixed station, of small interannual changes in meridional transport parameters. There is also some evidence which indicates that the polar stratosphere at high latitudes may represent an important ozone storage resevoir which tends to compensate for large scale changes observed in the regions outside of the polar stratosphere. Possible consequences of this are that the global trends derived from ground-based ozone measurements may not be valid and furthermore that the current satellite techniques by themselves may not be sufficient. An ozone monitoring system which includes observations from satellites, ground-based stations, balloons and rockets may be necessary.

Paper 29

VOL. 84, NO. C11 JOURNAL OF GEOPHYSICAL RESEARCH NOVEMBER 20, 1979

Seasonal and Interannual Variations in Total Ozone Revealed by the Nimbus 4 Backscattered Ultraviolet Experiment

ERNEST HILSENRATH AND DONALD F. HEATH

Laboratory for Atmospheric Science, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771

BARRY M. SCHLESINGER

Systems and Applied Sciences Corporation, Riverdale, Maryland 20840

The first 2 years of backscattered ultraviolet (BUV) ozone data from the Nimbus 4 spacecraft have been processed to a more refined level. The seasonal variations of total ozone for the period April 1970 to April 1972 are described using daily means for 10° latitude zones and a time-latitude cross section. In addition, the BUV data are compared with analyzed Dobson data and with infrared interferometer spectrometer data also obtained from the Nimbus 4 spacecraft. A harmonic analysis was performed on the daily zonal means. Amplitudes, days of peak ozone values, and percentage of variance have been computed for the annual and semiannual waves and for higher harmonics of an annual period for the 2 years. Asymmetries are found in the annual waves in the two hemispheres, with a subtle interannual difference which may be due to changes in the general circulation. A significant semiannual component is detected in the tropics for the first year. This component appears to result from influences of the annual waves in the two hemispheres. A search for shorter periods using the harmonic analysis revealed no periodicity whose amplitude was higher than the noise level.

INTRODUCTION

The total column amount of atmospheric ozone has been observed from the ground for several decades from stations sparsely scattered around the earth. These early observations were made because it seemed that total ozone acts as a tracer of lower stratospheric motions. Moreover, the vertical distribution of ozone is important to the radiative balance of the stratosphere because of its strong absorption in the ultraviolet and its emission in the infrared. Recent interest in stratospheric ozone has been aroused because of possible depletion by anthropogenic activity, which results from catalytic agents reaching the stratosphere. A columnar reduction in ozone will allow increased solar ultraviolet radiation to reach the earth's surface and would likely affect the biosphere. An indirect result of ozone depletion may be a change in the climate because of radiative coupling between the troposphere and the stratosphere.

A concise summary of the global variations in the total amount and vertical distribution of ozone is given by *Dütsch* [1974], derived mainly from ground-based data. Although a fairly accurate description of the seasonal and latitude variations is given, he feels that the ground observing network is not adequate for a synoptic representation of ozone for tracer studies or monitoring global ozone trends. *London et al.* [1976] have constructed hemispheric maps of monthly mean ozone, using all of the available total ozone data from the Dobson network from 1957 to 1967. These data cover nearly a decade and could be used for determining interannual variations, except in the southern hemisphere, where there are too few observing stations. Global ozone trends have been determined by *Angell and Korshover* [1978] from the ground-based observations and then compared to other geophysical phenomena which could affect the stratosphere. Ozone trends have been modeled by *Penner and Chang* [1978], who found the Angell and Korshover analysis to be consistent with modeled variations of solar flux due to the 11-year solar cycle. Multidimensional photochemical-dynamical models can now predict the seasonal global distribution of ozone. Although these models to some degree use basic dynamic variables, they are normally verified by comparing the model results to the observations.

The quality of the observations suffers because of their poor distribution over the globe. In addition, the ground stations have not been intercalibrated in the past. The backscattered ultraviolet experiment (BUV) on Nimbus 4 has provided ozone measurements with nearly global coverage and an accuracy comparable to that of a well-calibrated Dobson station. This paper will present an accurate description of the seasonal and interannual variations of total ozone over a 2-year period as derived from the satellite measurements. An accurate description of the global seasonal variations is important for detecting ozone trends and recognizing nonperiodic phenomena that may be associated with transient events, such as stratospheric warmings and geomagnetic activity. In addition, satellite data should be used to verify models, since these observations are more globally representative.

The analysis of the satellite ozone data performed in this paper is in terms of zonal means, since a time-varying globally distributed variable can be conveniently described in this manner. This is particularly appropriate for total ozone, since it is a strong function of season and latitude. The next section reviews the spacecraft measurement. This is followed by a discussion on the procedure used in forming the zonal means and the application of the harmonic analysis for describing the seasonal variations. The BUV data are then compared to similar data from the infrared interferometer spectrometer (Iris), also flown on Nimbus 4. They are also compared to the analysis of Dobson network data for the same period, presented in the form of a time-latitude cross section. The last section deals with a description of the seasonal, interannual, and shorter variations detected by the BUV as derived from the harmonic analysis of the zonal mean data.

Fig. 1a Coverage map for 1 week in April for Nimbus 4 BUV observations. Each spot indicates a square measurement sample of approximately 200 km on a side. The grey code ranges from 190 to 500 DU from dark to light. The continental outline would be the same as that shown in Figure 1b.

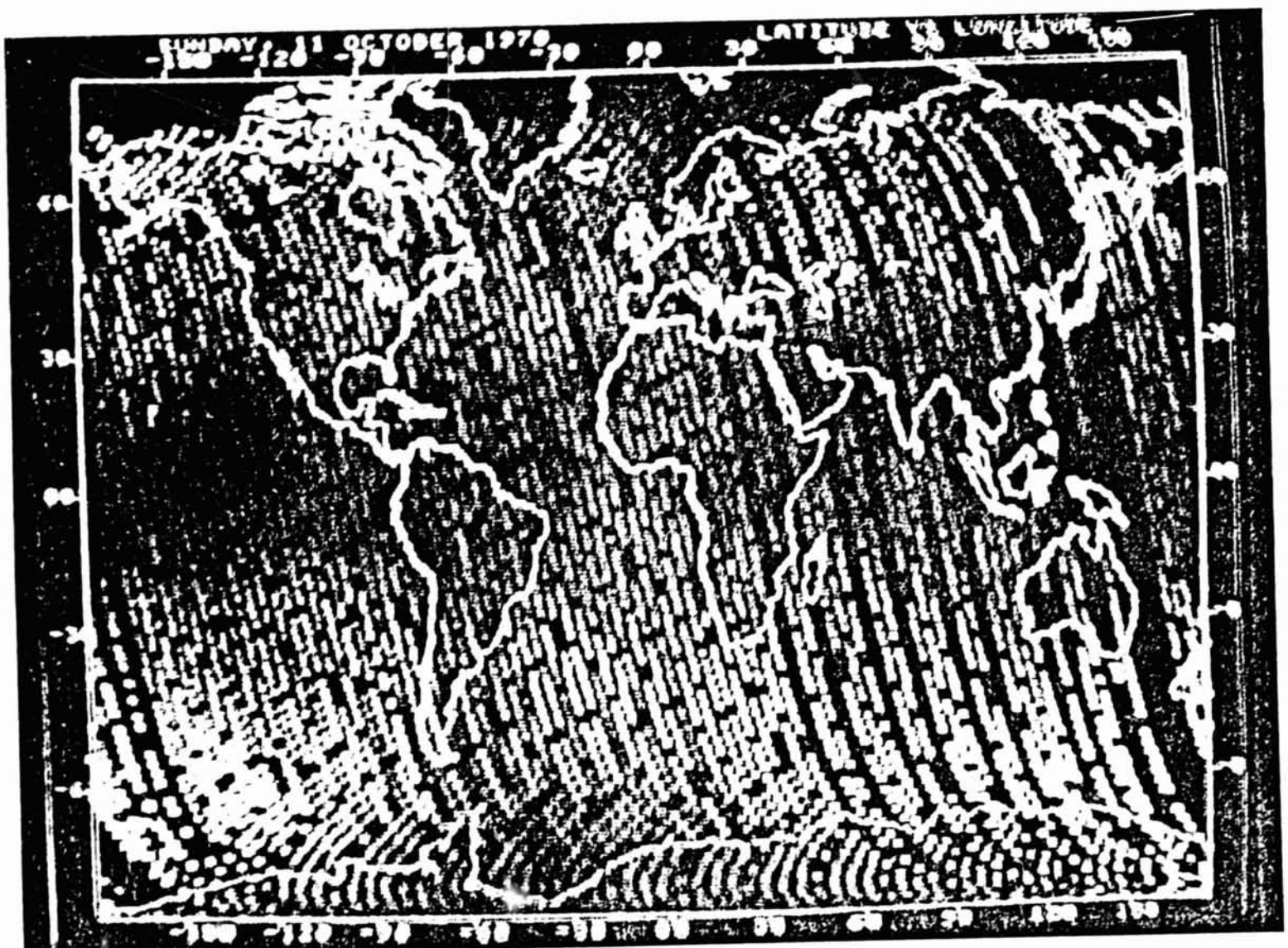


Fig. 1b Same as Figure 1a but for 1 week in October

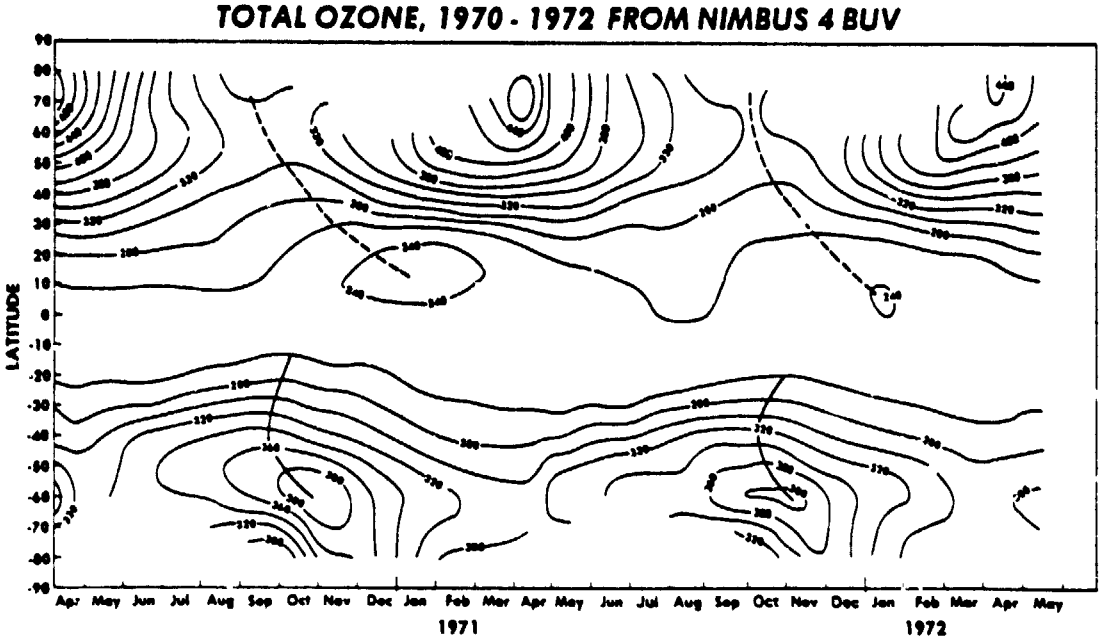


Fig. 2. Time-latitude cross section of total ozone for the period April 1970 to May 1972 derived from 10°-wide daily zonal means which are smoothed with a binomial filter. Contours are 20-DU increments.

Measurement

The Nimbus 4 spacecraft was launched April 8, 1970, carrying the BUV spectrometer for measurements of the columnar amount of ozone and the vertical distribution above 30 km. The BUV continued to operate until October 17, 1977, when it was turned off because of insufficient spacecraft power. Results based on the initial data were reported by *Krueger* [1974], *Ghazi* [1976], and others. These early analyses used existing instrument calibrations and orbital engineering data. At that time the satellite-measured total ozone values were adjusted to the Dobson measurements by using a linear regression relationship developed from direct comparison of BUV-measured values and Dobson values. Because of the recently recognized importance of the BUV data set a task was initiated to reprocess the data and to continue processing the available data until the time that the instrument was turned off. The reprocessing activity included the following procedures: establishment of a primary data base with screened, earth-located, and calibrated data for conversion to radiance values; improvement of the algorithms for processing the primary data base of radiance values to ozone values; and, finally, validation of the data to provide a consistent ozone data base. Details on satellite coverage, data quality checks, algorithm improvements, error analyses, and validation methods will be published elsewhere.

The comparability of BUV with the Dobson network has been extensively analyzed by *Miller et al.* [1978] and *Fleig et al.* [1978]. In a direct comparison of the AD wavelength pair, direct sun Dobson, and BUV over-flight values within 2° of the station, they obtained a correlation coefficient of 0.95 and an average difference of 12 Dobson units (DU), with Dobson higher for the first 2 years of operation. The change in the BUV-Dobson difference over the 2 years is about 2 DU, or less than 1%, indicating a high degree of stability in the spaceborne instrument relative to the ground-based network over the first 2 years. The comparability of BUV and the Dobson network from 1972 to 1977 is under investigation by the authors.

The first 2 years of total ozone data discussed here, covering the period April 1970 to April 1972, contain about 300,000 total ozone values per year. The Nimbus 4 orbital parameters and the BUV instrument description are detailed in the *Nimbus-4 Users Guide* [*Nimbus Project*, 1970]. There are about 13 orbits per day, with successive crossings of the equator separated by 27° in longitude. The total ozone measurement is limited to between 81° north and south and to a maximum solar zenith angle of 85.7°. The orbit is sun synchronized so that the northward equator crossing occurs near local noon. At high latitudes the observations approach the morning terminator in the northern hemisphere and the evening terminator in the southern hemisphere. Figures 1*a* and 1*b* are typical coverage maps for 1 week of data in April and October, where each spot indicates a measured ozone value. These figures will be referred to again.

The BUV measurement scheme has been previously described by *Mateer et al.* [1971] and is briefly summarized here. The instrument contains a monochromator, which sequentially scans 12 narrow wavelength bands in a region between 250.0 and 340.0 nm, and a photometer fixed at 380.0 nm. Both instruments measure the backscattered and reflected earth radiances in the nadir and have an instantaneous field of view of about 200 km on the earth's surface. Total ozone is derived from a measurement of the solar irradiance, the radiances at 312.5, 331.2, 317.5, and 339.8 nm backscattered in the troposphere which are attenuated by the ozone layer, and the effective surface reflectance determined from the photometer. The measured backscattered radiances are compared to those computed from 21 standard ozone profiles compiled from rocket and balloon soundings [*Hilsenrath et al.*, 1977]. A total ozone value can then be obtained from the table of pre-

computed values containing solar zenith angle dependence and reflectance.

Procedure

The latitude zones over which the average ozone is calculated are centered at the equator and are at 10° intervals toward the pole. Because of the high-latitude limit of the satellite the highest-latitude zones are for 75°–85° in each hemisphere. For each zone a daily mean is calculated by averaging all ozone values given by the satellite for the given latitude range. Typical standard deviations of these means range from 10 DU near the equator to 30 DU at high latitudes. These standard deviations reflect primarily the low measurement variability at low latitudes and the larger variations with longitude at high latitudes. Zonal means are based upon approximately 50 measurements per day.

For the two highest-latitude zones in each hemisphere, covering 65°–85° in latitude, ozone values are not available during winter, when the zone has a 24-hour night. Because a harmonic analysis cannot be carried out accurately when there are large gaps in the data, an estimate of the daily zonal means at the highest latitudes during this period is derived by extrapolation from lower-latitude zonal means. The extrapolation from lower latitudes is based on the behavior of high-latitude ozone during those periods when measurements are available. The extrapolation procedure is as follows: When ozone values Ω_{j-1} and Ω_j are available for zones $j-1$ and j, respectively, but no value is available for zone $j+1$, the estimated value for zone $j+1$ is given by

$$\Omega_{j+1} = \max \{0.96[\Omega_j + (\Omega_j - \Omega_{j-1})], 0.95\Omega_j\} \quad (1)$$

Because some ozone values are extrapolations, results for the zones centered at 70° and 80° should therefore be considered estimates.

The daily zonal means for each of the 17 latitude zones have been converted into time series, and a harmonic analysis has been conducted separately for each. Amplitudes, days of peak ozone, and the percentage of variance have been computed for each harmonic. The first 50 harmonics of a 1-year (365.24-day) period have been computed for the 365-day periods April 10, 1970, to April 9, 1971, and April 10, 1971 to April 8, 1972. Also, the first 50 harmonics of a 2-year (730.48-day) period have been computed for the interval April 10, 1970 to April 8, 1972. These are the first 730 days for which data are available from Nimbus 4 BUV.

In a harmonic analysis the ozone $\Omega(t)$ is expressed in the following form:

$$\Omega(t) = \langle\Omega\rangle + \sum_m C_m \cos\left(\frac{2\pi mt}{T} + \phi_m\right) \quad (2)$$

where $\langle\Omega\rangle$ represents the mean value, C_m the amplitude for harmonic m, ϕ_m its phase, and T the period corresponding to the first (fundamental) harmonic. The peak day, or the day when ozone is a maximum for each latitude zone, is determined from the maximum value of the cosine function in (2).

The percent of the variance attributable to harmonic m is given by $C_m^2/2\sigma^2$, where σ^2 is the variance of the zonal mean over all days in the year. Since occasional days are missing in the data, values are derived for these days by linear interpolation when calculating the average ozone $\langle\Omega\rangle$. The uncertainties of the amplitudes and peak days have been estimated by calculating harmonics for 1-year periods with different starting days. The derived uncertainties in the annual and semiannual are approximately 10% of the amplitude and 3–4 days for the peak time and are not significant for this study.

Results

Time-Latitude Cross Sections

A time-latitude cross section, depicted in Figure 2, was generated from smoothed averaged daily means for each latitude zone. Total ozone is shown as contours with 20-DU increments for the 26-month period (missing contours represent the polar night). The seasonal march of total ozone is evident in the two hemispheres. Ozone has a spring high and a fall low at middle to high latitudes in both years. However, there are distinct differences in the seasonal trends for each hemisphere. In the northern hemisphere the spring maximum occurs nearly simultaneously at middle and high latitudes in late March. The maximum values occur near the pole, with total ozone amounts of about 500 DU in the spring of 1970 decreasing to 480 and 460 DU in the springs of 1971 and 1972, respectively. This is consistent with the analysis of *Angell and Korshover* [1978], who show from Dobson data a comparable decrease in the north temperate and polar regions for this same time period.

In the southern hemisphere, on the other hand, the spring maximum occurs first in September at 50°S, where ozone values are substantially lower than those in the northern hemisphere spring. The spring maximum values occur about 1 month later in the polar regions. The asymmetry in the winter buildup of the two hemispheres has been related to the well-known differences in the upper tropospheric and lower stratospheric circulation features in the two hemispheres [*Newell et al.*, 1974; *Dütsch*, 1974]. In the northern hemisphere, eddy processes associated with the intense winter planetary waves transport ozone poleward from mid-latitudes. In the southern hemisphere, however, the circulation is generally more zonal, and there is a northward component of eddy fluxes in the south polar regions. Furthermore, poleward transport of ozone may be further delayed because stratospheric warmings may be less intense or because of the resistance to breakdown of the southern hemisphere polar vortex as noted by *McIntyre* [1978].

In the northern hemisphere the strongest ozone latitudinal gradient can be associated with the planetary or synoptic scale waves and the jet stream. This is clearly illustrated in Figure 1*a*, which shows largest ozone values at the time of the northern hemisphere spring maximum. However, in the southern hemisphere the gradients are weaker, and the planetary waves, on a weekly average, do not appear. This is illustrated in Figure 1*b*, which is a coverage map for 1 week in October, the time for southern hemisphere spring. Note the low ozone values at the south pole near 30°E. *Miller et al.* [1978] have found a strong negative correlation of total ozone departures from the zonal mean with departures of the height of the 100-mbar level from its zonal mean in the same zone. Their analysis provides additional evidence of the importance of synoptic scale waves in transporting ozone poleward.

In the tropical regions the total ozone amount and the seasonal variations are considerably smaller than at higher latitudes. The yearly average value is about 250 DU. The ozone minimum is centered below the equator in April and moves northward as the year progresses. This cycle is repeated in the second year. Because the sun-earth distances and the solar

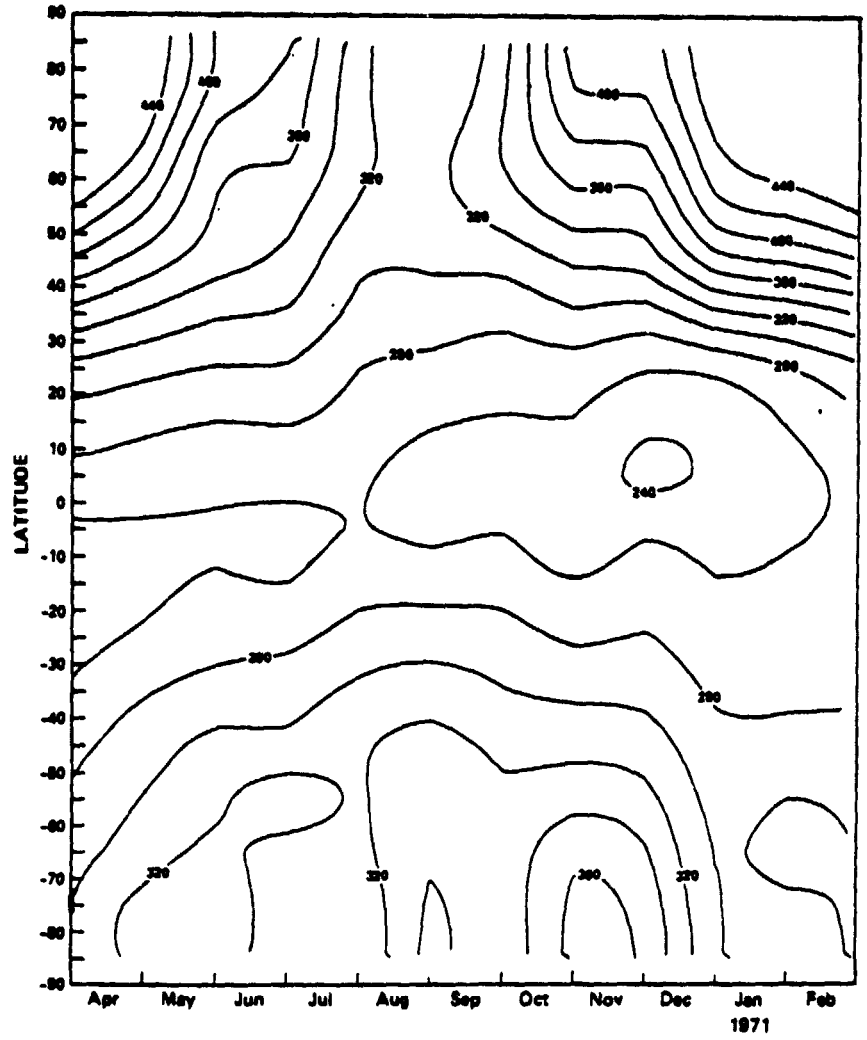


Fig. 3. Time-latitude cross section derived from analysis of Dobson data using monthly zonal averages (J. London, private communication, 1979).

declination vary over the year, this effect could be interpreted as a photochemical response, since the tropics are the source region for ozone. However, the effect appears rather to be the result of the seasonal variation in the two hemispheres and will be discussed in more detail later. Also, note that the ozone minimum in the northern tropics, which appears just after the southern hemisphere spring maximum and at the beginning of the northern hemisphere winter buildup, occurs in both years.

The global mean ozone trend can be easily derived by summing the area weight of the zonal mean ozone values for each day. For each of the 2 years studied here, this calculation shows minimum global mean ozone in October and November, just after the southern hemisphere spring maximum and prior to the northern hemisphere winter buildup. The amplitude of the annual oscillation is about 20 DU, about 5 times higher than that calculated by *Keating* [1978], who used the Nimbus 4 Iris total ozone data. The BUV data, however, are consistent with the mean global trend calculated from J. London's (private communication, 1979) analysis for the same time period. For the 2-year time period the global mean ozone trend derived from BUV shows a gradual decrease which is consistent with the analysis of *Angell and Korshover* [1978], who use Dobson data. Trends and variations in global mean ozone measured from satellites and compared with ground observations will be discussed in detail by D. F. Heath et al. (in preparation, 1979) and will not be pursued further here.

Comparison With Dobson Analysis

The Dobson network of stations provides an alternate source of total ozone values. *London et al.* [1976] first derived monthly hemispheric ozone maps from the Dobson data for the period 1957–1967. J. London (private communication, 1979) has provided additional maps for the period 1968–1975. Data for 1970–1971 are used to form a time-latitude cross section (Figure 3) similar to that shown in Figure 2 for BUV data. In the northern hemisphere the seasonal features shown by Dobson and BUV are nearly the same. The general pattern of ozone values and the times of the spring maximum and fall minimum agree reasonably well but with some differences in detail. In the tropics both data sets show the ozone minimum moving from south of the equator in April to its most northerly point in January. In the southern hemisphere there are significant differences. The spring buildup at middle to high latitudes that appears in the BUV cross section does not ap-

pear at all in the analyzed Dobson cross section. This difference may be in part due to the scarcity of ground observing stations in the southern hemisphere (eight stations south of 30°S reported in 1970 and only two reported in October).

In order to understand further the discrepancy between the two cross sections the BUV data can be examined in the form of latitude-longitude maps. Figure 4 is a map of contoured ozone values at 20-DU increments for October 1970. A very strong wave one type feature appears at about 60°S, with the highest ozone values southeast of Australia. This feature has been detected in the past from Dobson [*London et al.*, 1976], although the ozone values were considerably lower than the 480 DU measured by the BUV. It is concluded then that the differences in the southern hemisphere illustrated in Figures 2 and 3 result mainly from the absence of Dobson stations, which prevented detection of the large ozone high by the Dobson analysis. The discrepancy recurs when comparing the BUV and Dobson analyses for the period March 1971 to March 1972. It should be noted, however, that the seasonal trends for the 2 years shown in Figure 2 agree fairly well with the 10-year averaged cross section derived by *London et al.* [1976]. This would imply that the BUV data for the 2 years studied here more closely describe the mean global ozone.

Comparison With Iris

The Iris was also flown on Nimbus 4 and measured total ozone at the same time as BUV. An analysis of the derived temperature and ozone fields was reported in detail by *Prabhakara et al.* [1976]. *Prior and Oza* [1978] performed a preliminary comparison of BUV, Dobson, and Iris data and showed that the BUV and Dobson agree better in general with each other than with Iris. However, Figure 4 can also be directly compared with the October analysis (Figure 13 in the Prabhakara et al. paper) derived from Iris. The ozone high over Manchuria, the low over the North Atlantic, and the closed contour of 240 DU in the vicinity of the Indonesian islands also appear in their analysis. The ozone high in the southern hemisphere high latitudes centered at 170°E longitude also appears in the Iris data, but this feature is about 20 DU lower. However, when correcting the Iris total ozone measurement underestimate of spring ozone values and an over-estimate in fall [*Prabhakara et al.*, 1976], the monthly analyses from the two instruments are nearly identical.

Harmonic Analysis

The seasonal variations and the year to year differences in the global total ozone were obtained from a harmonic analysis of the daily zonal means by using the procedures discussed in a previous section. Figure 5 illustrates daily zonal means for five latitude zones for the period April 10, 1970, to April 9, 1971. The daily zonal means were normalized and smoothed, in this example only, by setting the maximum and minimum values to +1 and −1, respectively, and smoothing with a binomial filter with a 10-day half width. At all latitude zones except the equator an annual wave is clearly present. Periods of shorter variability are also evident. For some zones, for ex-

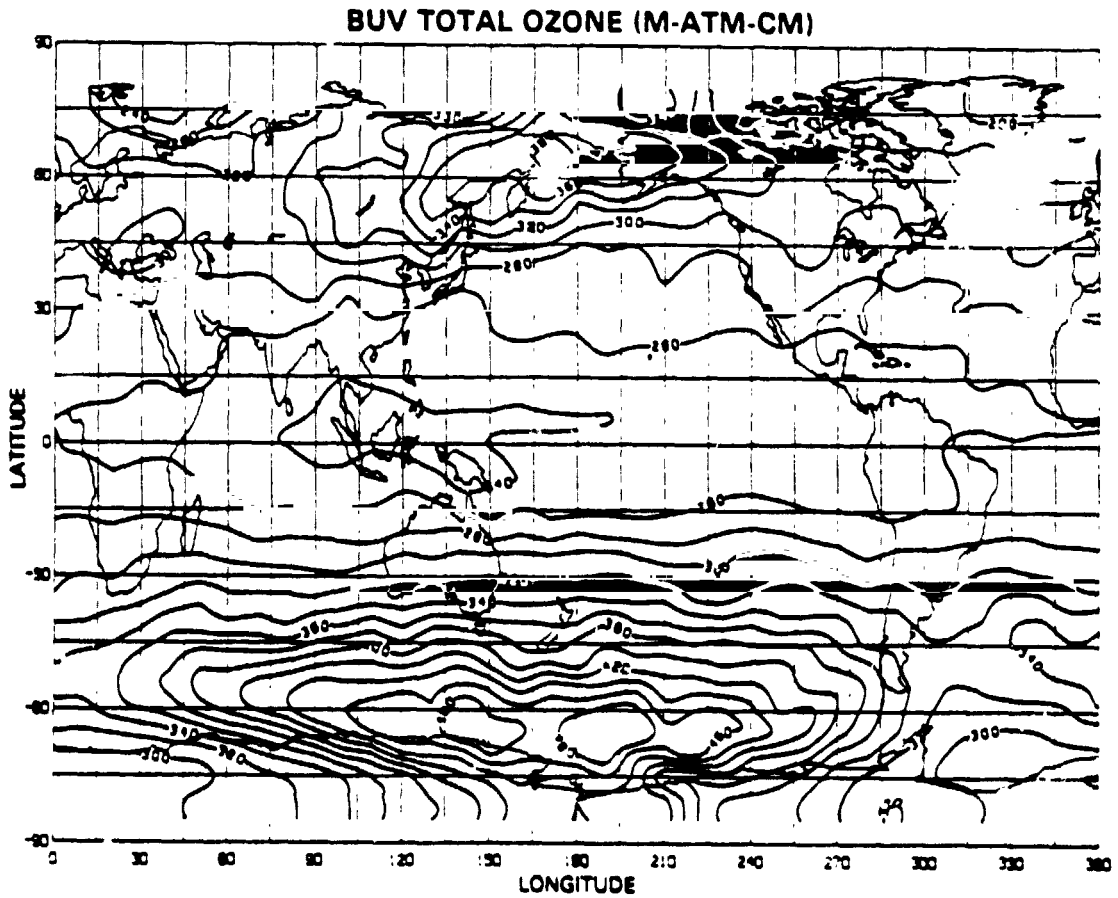


Fig. 4. Average total ozone for October 1970 contoured in 20-DU increments.

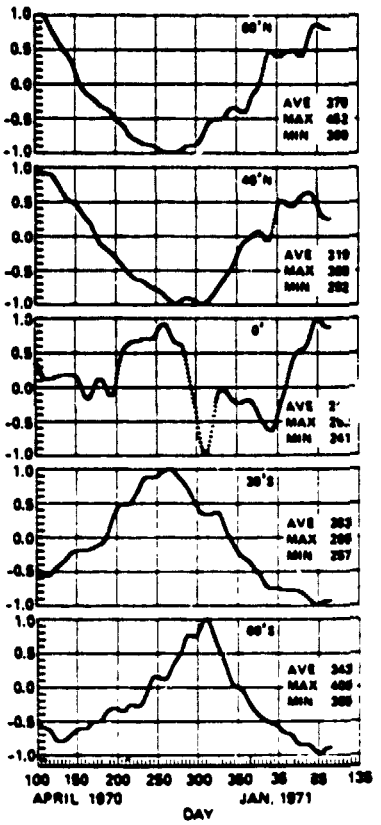


Fig. 5. Normalized and smoothed (see text) daily zonal means illustrating periodic behavior for five latitude zones. The abscissa is day of the year number. The average, maximum, and minimum total ozone for the year in each zone also is indicated.

ample, 30°S, the variations are nearly sinusoidal, while at 60°S, for example, the curve differs significantly from a pure sinusoid. In addition, periodicities of the order of 2-3 weeks appear during the southern hemisphere winter buildup. The phase shift with latitude for the time of the maximum value in southern hemisphere zones, discussed earlier, is clearly evident in this figure.

The annual wave or first harmonic in the Fourier expansion is shown in Figure 6. The amplitude of and percent of variance in the first harmonic are plotted as functions of latitude. Both years shown in the single figure permit a direct comparison. It should be clear that the data shown in this figure (and subsequent figures) are representative of information in Figure 2. The important features are as follows:

1. In the northern hemisphere the amplitude of the annual wave increases with latitude. In addition, the first harmonic contributes more than 90% of the variance from subtropical latitudes to the pole. These features are essentially the same for the 2 years.

2. In the tropics the amplitude of the annual wave is a minimum and is only a few Dobson units. The percent of the variance in the first harmonic is nearly zero at the equator, at least for the first year. The location of the minimum amplitude shifts about 10° southward in the second year.

3. In the southern hemisphere the amplitude of the annual wave is greatest at mid-latitudes and decreases toward the south pole. The percent of the variance in the annual wave decreases near the pole, indicating that the total ozone annual wave is less pronounced in the south polar region than in the north. At mid-latitudes there is a 10° southward shift in the location of the maximum amplitude in the second year, comparable to that detected in the tropics. These shifts could result from year to year changes in the strength and location of the Hadley cell circulation in the southern hemisphere. Interannual changes in the strength of the Hadley cell determined from a study of 10 years of northern hemisphere wind data have been discussed by *Rosen and Wu* [1976]. Moreover, *Newell and Wu* [1978] found a correlation between total ozone from long-term monthly mean Dobson observations and geopotential thicknesses for the same period. They concluded that circulation changes due to variations in the Hadley cell govern the year to year changes in ozone. Another possible cause for this southward shift may be due to the quasi-biennial oscillation (QBO) of tropical stratospheric winds. A QBO has been detected in Dobson total ozone data by *Angell and Korshover* [1978]. A strong QBO in total ozone has also been detected at low latitudes in a preliminary analysis of the remaining 5 years of BUV data. The effect of the QBO on poleward transport of ozone, the annual waves in the two hemispheres, and the longer-term ozone trend is under study.

Information on the semiannual component in the seasonal march of ozone is contained in the second harmonic of the Fourier expansion of the daily zonal means. A semiannual component may be expected, since the sun crosses the equator twice and a photochemical response may be detected in the tropics. On the other hand, ozone correlations with the semiannual oscillation in the temperature and winds detected in the tropics in the upper stratosphere [*van Loon et al.*, 1972] or with the winter high-latitude/tropical disturbances discussed by *Fritz and Soules* [1972] are also possible. These phenomena are of great interest but require a more detailed analysis than that performed to show correlations with ozone. Instead, the semiannual component detected in the tropics from the satellite measurement for the 2 years studied here will be shown to have a different origin. Figure 7 presents the semiannual component in the same format as Figure 6, except that the ordinates for both the amplitude and the percent variance are half those of that figure. The important features are as follows:

1. At mid-latitudes for both years the semiannual wave is small and of the order of its uncertainty. Clearly, the annual wave dominates in this latitude region. In northern high latitudes the amplitude increases; however, the variance is near the noise level.

2. In the tropics the semiannual component becomes significant in the first year but is less so in the second year. From 0° to 10°S the amplitude of the semiannual component is comparable to that of the annual component and contains about 40% of the variance for the first year.

3. At high latitudes in the southern hemisphere the semiannual oscillation may also be important. The amplitude appears comparable to that of the annual wave, and the variance of the semiannual wave is 3 times larger than that of the annual wave. However, the significance of this feature is uncertain, since nearly one half of the values appearing in the highest-latitude zones (above 70° in winter) are the result of extrapolation from lower latitudes, using the technique described earlier. It is interesting to note, however, that the *London et al.* [1976] 10-year average of analyzed Dobson data gives a very similar result, even though there are no ground observations in the winter. *Wilcox et al.* [1977] have also detected a semiannual component in the vertical distribution of ozone from balloon measurements in the northern hemi-

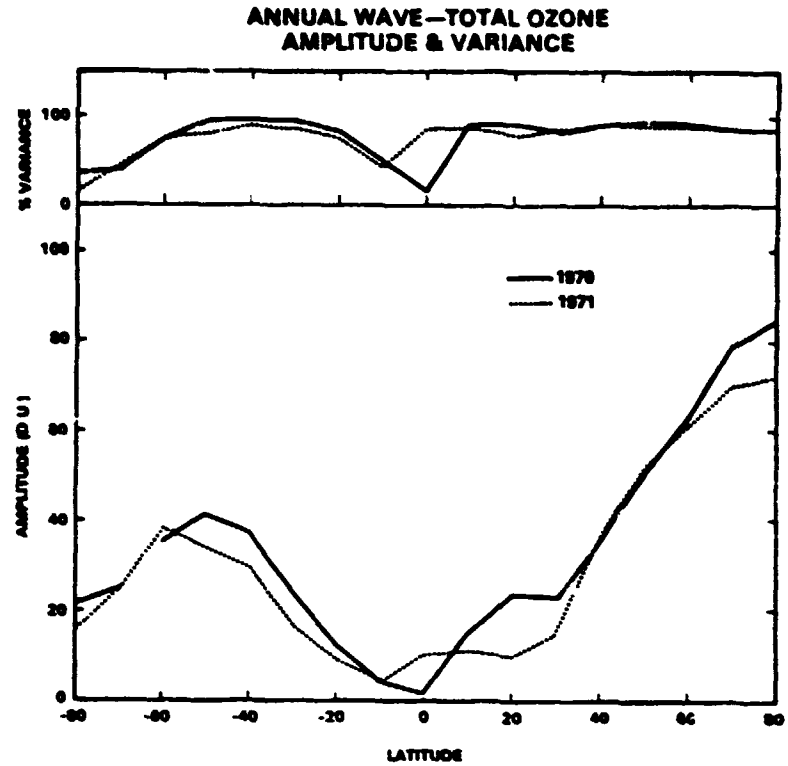


Fig. 6. (Bottom) Amplitude of annual wave and (top) variance of first harmonic derived from daily zonal means for the first and second year of BUV total ozone data.

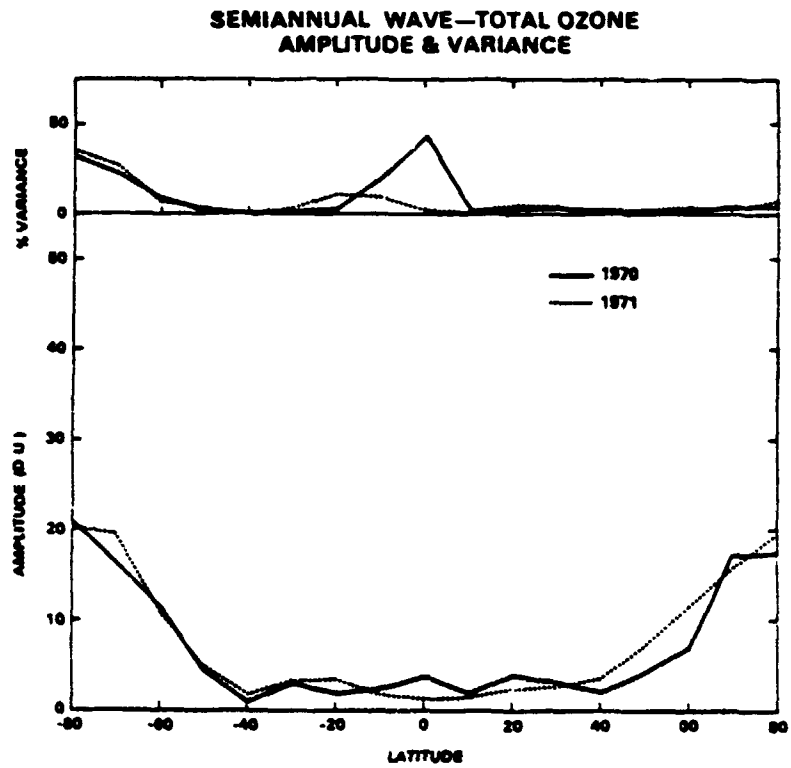


Fig. 7. (Bottom) Amplitude of semiannual wave and (top) variance of second harmonic derived from the daily zonal means for the first and second year of BUV total ozone data.

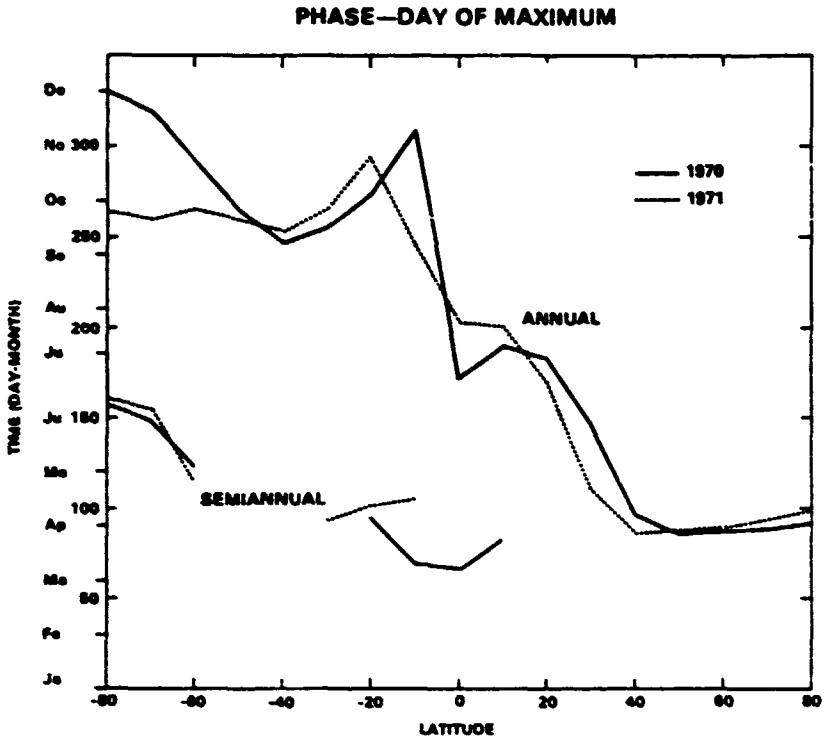


Fig. 8. Time of maximum ozone value derived from phase in the harmonic analysis of daily zonal means of BUV data.

sphere. (They had no data in the southern hemisphere.) Finally, *van Loon et al.* [1972] found that the amplitude of the semiannual wave in the zonally averaged temperatures in the lower stratosphere (50–10 mbar) is a maximum at high latitudes in the southern hemisphere. In the tropics and the northern hemisphere this maximum occurs above 5 mbar and should have little effect on the total ozone. Therefore BUV data are consistent with these results; however, a more detailed study is required to reveal the mechanism linking the relationship between the temperature, the winds, and the total ozone, revealed in the semiannual wave.

The harmonic analysis also provides the time of the maximum ozone value derived from the phase and period in (2) in each latitude zone and for each harmonic component. Figure 8 illustrates the time of the maximum ozone value as a function of latitude for both the annual and the semiannual waves for each year. The ordinate is given by month as well as day of the year. The significant aspects of the phase are summarized next. The conclusions are consistent with the interpretation of the time-latitude cross section shown in Figure 2.

In the northern hemisphere the maximum of the annual wave occurs in late winter from middle to high latitudes within only a few days. This could be due to the anticipated strong eddy motions at that time of year, but the detailed mechanisms causing the maximum values to occur in so short a period of time are not clear. After passing through a 6-month phase shift at the equator the southern hemisphere spring maximum occurs first at 40°S in September and then spreads poleward and equatorward, reaching maximum values in these regions in late November.

For the semiannual wave the time of the first maximum was plotted only at latitudes where the percent of the variance in the second harmonic is greater than 4%. At low latitudes the first peak occurs about April, coincident with the northern hemisphere spring maximum of the annual wave. The second maximum of the semiannual component occurs 6 months later, in October, coincident with the maximum of the annual wave in the southern hemisphere subtropics. Close inspection of Figure 5 supports this result. The maximum annual ozone value at 30°S occurs on day 265, while the maximum annual value at 40°N occurs about day 80, in 1971, which very nearly corresponds to peak days at 0°S. This result does not exclude a possible response to the solar equatorial crossing, but it seems more likely that the semiannual wave detected at low latitudes in the first year is the result of the annual wave in the two hemispheres, where the phases are 6 months apart.

Shorter Periods

The results of the harmonic analysis were examined for shorter periodicities. For periods of 1 month or shorter the noise level ranges from 2 DU at low latitudes to 5 DU at high latitudes. No periodicity with higher amplitude was found, although variations of 2–3 weeks are evident in Figure 5, particularly during the first 6 months at 60°S. Harmonic analysis is not a particularly sensitive method for determining shorter periods. Further work in this area, using more sophisticated techniques, is under way.

Conclusion and Summary

The refined BUV data from Nimbus 4 have been analyzed here for the period April 1970 to May 1972. The coverage and overall sensitivity of the satellite measurement permit an accurate determination of the seasonal march of total ozone in the two hemispheres previously unobtainable from the ground-based observations, especially in the southern hemisphere. This study generally confirms the seasonal variations determined from several decades of Dobson data. With only 2 years of satellite data some interesting interannual differences

are revealed. Comparison of time-latitude cross sections derived from BUV and analyzed Dobson data shows significant differences in the southern hemisphere for the years 1970–1972. The spring buildup detected by the BUV at mid-latitudes does not appear at all in the analyzed Dobson data. This discrepancy is most likely due to the scarcity of Dobson stations in this half of the globe.

A harmonic analysis of the daily zonal means shows asymmetries between the annual waves in the two hemispheres. In the northern hemisphere the amplitude of the annual wave increases when going toward the pole, and the maximum value occurs nearly simultaneously at middle to high latitudes. In the southern hemisphere the maximum amplitude of the annual wave occurs at about 50°S. Maximum amplitude occurs first also near this latitude, and it then spreads equatorward and poleward. A semiannual component was detected at the equator in the first year and can be explained as the result of the annual waves in the two hemispheres whose phases are 6 months apart. The semiannual wave detected in the southern hemisphere high latitudes for both years can be correlated with other observations but cannot be satisfactorily explained. However, a similar feature is observed in the *London et al.* [1976] Dobson analysis averaged over 10 years. A search for shorter periods (higher-order harmonics) using a harmonic analysis with a 1-year period revealed no periodicity whose amplitude was significantly higher than the noise level.

A 10° southward shift in the location of the annual wave in the southern hemisphere was detected for 1971 as compared to 1970. This may be the result of interannual variations in the Hadley cell circulation which may have caused the decrease in global mean ozone values reported for this period. An alternative explanation for the shift may be the quasi-biennial oscillation of stratospheric winds at low latitudes. This oscillation has been detected in a preliminary analysis of 7 years of BUV data.

Work is under way to analyze the remaining 5 years of BUV data as they become available. Of particular importance is the development and testing of analysis schemes which can handle missing data, because the BUV operation was considerably reduced after the second year.

Acknowledgments. The authors would like to acknowledge the remainder of the Ozone Processing Team, which was directed by Albert Fleig at Goddard Space Flight Center. Special appreciation is given to James Gatlin, who retraced the BUV instrument performance 7 years after it was flown and the production of the coverage maps shown here, and to V. G. Kaveeshwar, who organized and implemented the data processing schemes used to produce the data. We also thank Julius London for allowing us to use the analyzed Dobson data prior to their publication.

References

Angell, J. K., and J. Korshover, Global ozone variations: An update into 1976, *Mon. Weather Rev.*, *106*, 725–737, 1978.

Dutsch, H. U., The ozone distribution in the atmosphere, *Can. J. Chem.*, *52*(8), 1491–1504, 1974.

Fleig, A. J., R. S. Fraser, B. W. Guenther, D. F. Heath, E. Hilsenrath, L. V. Novak, V. G. Kaveeshwar, R. D. McPeters, C. L. Mateer, and A. J. Miller, Global ozone determination from Nimbus-4 BUV spacecraft data, in *Developments in Atmospheric Science, 9, Remote Sensing of the Atmosphere*, edited by A. L. Fymat and V. E. Zuev, pp. 97–106, Elsevier, New York, 1978.

Fritz, S., and S. D. Soules, Planetary variations of stratospheric temperatures, *Mon. Weather Rev.*, *100*, 582–589, 1972.

Ghazi, A., A. Ebel, and D. F. Heath, A study of satellite observations of ozone and stratospheric temperature during 1970–1971, *J. Geophys. Res.*, *81*(30), 5365–5376, 1976.

Hilsenrath, E., P. J. Dunn, and C. L. Mateer, Standard ozone profiles from balloon and rocket data for satellite and theoretical model input, in *Collection of Extended Summaries of Contributions Presented at the Joint Assembly CMUA Sessions IAGA/IAMAP, Seattle, Washington, August 1977*, National Center for Atmospheric Research, Boulder, Colo., 41-1–41-6, 1977.

Keating, G. M., Relationship between monthly variations of global ozone and solar activity, *Nature*, *274*, 873–874, 1978.

Krueger, A. J., Behavior of atmospheric ozone determined from Nimbus satellite backscattered ultraviolet data, *NASA/GSFC X Doc.*, *910-74-101*, 1974.

London, J., R. D. Bojkov, S. Oltmans, and J. F. Kelly, Atlas of the global distribution of total ozone July 1957–July 1967, *Tech. Note NCAR/TN/113+STR*, Nat. Center for Atmos. Res., Boulder, Colo., Jan. 1976.

Mateer, C. L., D. F. Heath, and A. J. Krueger, Estimation of total ozone from satellite measurements of backscattered ultraviolet radiances, *J. Atmos. Sci.*, *28*, 1307–1311, 1971.

McInturff, R. M. (Ed.), Stratospheric warmings: Synoptic, dynamic and general circulation aspects, *NASA Ref. Publ.*, *1017*, 1978.

Miller, A. J., B. Korty, E. Hilsenrath, A. J. Fleig, and D. F. Heath, Verification of Nimbus-4 BUV ozone data and requirements for operational satellite monitoring, Symposium on the Geophysical Aspects and Consequences of Changes in the Composition of the Stratosphere, Toronto, Canada, June 26–30, 1978, *Pap. 511*, World Meteorol. Organ., Geneva, 1978.

Newell, R. E., and M. Wu, A pilot study of concomitant changes in total ozone and atmospheric general circulation, *Quart. J. Roy. Meteorol. Soc.*, *104*(442), 999–1003, 1978.

Newell, R. E., J. W. Kidson, D. G. Vincent, and G. J. Boer, *The General Circulation of the Tropical Atmosphere, and Interactions With Extratropical Latitudes*, vol. II, MIT Press, Cambridge, Mass., 1974.

Nimbus Project, *The Nimbus-4 Users Guide*, 214 pp., National Space Science Data Center, Greenbelt, Md., 1970.

Penner, J. E., and J. S. Chang, Possible variations in atmospheric ozone related to the 11-year solar cycle, *Geophys. Res. Lett.*, *5*(10), 817–820, 1978.

Prabhakara, C., E. B. Rodgers, B. J. Conrath, R. A. Hanel, and V. G. Kunde, The Nimbus 4 infrared spectroscopy experiment, 3, Observations of the lower thermal structure and total ozone, *J. Geophys. Res.*, *81*, 36, 1976.

Prior, E. J., and B. J. Oza, First comparison of simultaneous Iris, BUV and ground-based measurements of total ozone, *Geophys. Res. Lett.*, *5*, 547–550, 1978.

Rosen, R. D., and M. Wu, Observational study of the interannual variability in certain features of the general circulation, *J. Geophys. Res.*, *81*, 6383–6389, 1976.

van Loon, H., K. Labitzke, and R. L. Jenne, Half yearly wave in the stratosphere, *J. Geophys. Res.*, *77*(21), 3846–3855, 1972.

Wilcox, R. W., G. D. Nastrom, and A. D. Belmont, Periodic variations of total ozone and its vertical distribution, *J. Appl. Meteorol.*, *16*(3), 290–298, 1977.

(Received April 20, 1979;
revised June 25, 1979,
accepted June 29, 1979)

Phil. Trans. R. Soc. Lond. A **296**, 191–204 (1980)
Printed in Great Britain

Satellite ozone measurements

BY A. J. KRUEGER, B. GUENTHER, A. J. FLEIG, D. F. HEATH, E. HILSENRATH, R. MCPETERS AND C. PRABHAKARA
Laboratory for Atmospheric Sciences, Goddard Space Flight Center, Greenbelt, Maryland 20771 *U.S.A.*

Three classes of ozone sounders have been developed since the first Echo Satellite measurements in 1960. They are the (1) backscatter ultraviolet (b.u.v.), (2) infrared limb and nadir radiance, and (3) stellar and solar occultation methods. With these techniques, ozone has been measured from 20 to 100 km. Tropospheric ozone measurements are beyond present technology, but total ozone is determined with the b.u.v. and nadir infrared methods.

Results from the occultation methods are excessively variable, indicating a need for refinement of the technique. Evidence from this technique for a mesospheric ozone maximum is strong. The most extensive set of ozone observations has come from the Nimbus 4 b.u.v. between 1970 and 1977. From this and earlier b.u.v. experiments, it has been determined that the ozone tends toward photochemical steady state densities for levels above 5 mbar and that the temperature coefficient near 1 mbar is approximately 1000 K, in accordance with odd hydrogen catalysis of ozone. At lower altitudes, odd nitrogen catalytic chemistry has been verified from b.u.v. ozone observations during a solar proton event. Ozone profile results from the limb infrared radiance method are discussed in a companion paper by Dr J. C. Gille.

Total ozone soundings from the b.u.v. agree with Dobson results and details of seasonal and latitudinal variations are now available. The total ozone field has also been inferred from nadir infrared radiance data. Monthly mean total ozone maps from this method and the b.u.v. each indicate standing waves at all latitudes.

INTRODUCTION

In the last 15 years, interest in stratospheric ozone has changed from scientific curiosity to the present need for monitoring for environmental reasons. Fortunately, a variety of remote ozone sounding techniques were proposed and tested with satellites during the 1960s and early 1970s so that the current needs for research and climatological data can be met with a combination of satellite and conventional sounding devices. This paper is a review of satellite methods and, in conjunction with Dr Gille's paper, a summary of principal findings.

Ozone molecules absorb radiation strongly in the middle ultraviolet (u.v.) Hartley–Huggins bands (220–320 nm), moderately in the infrared (9.6 μm), and weakly in the visible Chappuis bands (0.6 μm). Perhaps the first measurement of ozone to use artificial Earth satellites in connection with the selective absorption of light employed the Echo balloon satellite. Sunlight in the Chappuis band wavelengths which passed through the Earth's limb was measured with a ground based telescope after reflexion from the satellite (Venkateswaran *et al.* 1961). A useful remote sounding technique, namely *limb occultation*, was demonstrated, although, because of geometric problems, the resulting ozone profiles were incorrect. This technique, as with most others, had a pre-satellite origin; that of lunar eclipse photometry. A chronology of developments in limb occultation methods and in the other major satellite techniques for ozone measurement is given in table 1.

The second major remote sensing method to be proposed was the *backscatter ultraviolet* (b.u.v.) technique. Sunlight in the Hartley–Huggins bands penetrates to various stratospheric altitudes (which depend on the wavelength-dependent ozone absorption cross section) where the photons are either absorbed or Rayleigh scattered. Measurements of the Earth albedo at several u.v. wavelengths can yield an ozone profile. The method was proposed by Singer & Wentworth (1957) and newer methods for inversion of the albedo measurements have been developed and reviewed by Mateer (1972). An important extension of the b.u.v. method was to measure total ozone as suggested by Dave & Mateer (1967).

TABLE 1. SATELLITE EXPERIMENTS TO MEASURE OZONE

type	satellite	wavelengths nm	latitude coverage	comments	references
occultation					
solar	Echo 1	590, 529.5	17 N	Dec. 1960	Venkateswaran *et al.* (1961)
	U.S.A.F. 1962	260	33S–13S	July 1962	Rawcliffe *et al.* (1963)
	Ariel 2	200–400	50 S–50 N	April, May, Aug. 1964	Miller & Stewart (1965)
	AE-5	255.5	5 N	Dec. 1976	Guenther *et al.* (1977)
stellar	OAO-2	250	16 S–43 N	Jan. 1970, Aug. 1971	Hays & Roble (1973)
	OAO-3	258–343	12 S–3 N	July 1975	Riegler *et al.* (1976)
backscatter u.v.					
profile	U.S.A.F. 1965	284	60 S–60 N	Feb.–Mar. 1965	Rawcliffe & Elliott (1966)
	U.S.S.R.	225–307	60 S–60 N	Apr. 1965	Iozenas *et al.* (1969)
		250–330	60 S–60 N	June 1966	Iozenas *et al.* (1969)
	1966-111 B	175–310	80 S–80 N	1966	Elliott *et al.* (1967)
	OGO-4	110–340	80 S–80 N	Sept. 1967–Jan. 1969	Anderson *et al.* (1969)
	Nimbus 4 b.u.v.	255.5–305.8	80 S–80 N	Apr. 1970–July 1977	Heath *et al.* (1973)
	AE-5 b.u.v.	255.5–305.8	20 S–20 N	Nov. 1975–Apr. 1977	Frederick *et al.* (1977*a*)
	Nimbus 7 s.b.u.v.	255.5–305.8	80 S–80 N	Nov. 1978*	Heath *et al.* (1975)
total	Nimbus 4 b.u.v.	312.5–339.8	80 S–80 N	Apr. 1970–July 1977	Mateer *et al.* (1971)
	AE-5 b.u.v.	312.5–339.8	20 S–20 N	Nov. 1974–Apr. 1976	
	Nimbus 7 t.o.m.s.	312.5–339.8	global	Nov. 1978*	Heath *et al.* (1975)
infrared emission		μm			
profile	Nimbus 6 l.r.i.r.	9.6	65 S–90 N	June 1975–Jan. 1976	Gille (1979, this symposium)
	Nimbus 7 l.i.m.s.	9.6	65 S–90 N	Oct. 1978–May 1979	Nimbus project (1978)
total	Nimbus 3 i.r.i.s.	9–10 spectral scan	80 S–80 N		Hanel *et al.* (1970)
	Nimbus 4 i.r.i.s.	9–10 spectral scan	80 S–80 N	Apr. 1970–Jan. 1971	Prabhakara *et al.* (1976)
	Block 5 m.f.r. (4 flights)		global	Mar. 1977*	Lovill *et al.* (1978)
	Tiros *N* h.i.r.s.	9.71	global	Nov. 1978*	

* Currently in operation (Nov. 1979).

Satellite measurements of ozone profiles by the backscatter method were first made by Iozenas *et al.* (1969) and Elliott *et al.* (1967), who used wavelength scanning monochromators

viewing the nadir atmosphere. These were followed by a similar experiment on the OGO-4 satellite in 1967 (Anderson *et al.* 1969), in which data were collected over a 17-month period. The first major backscatter u.v. experiment with an instrument dedicated to ozone measurements was the b.u.v. on Nimbus 4 in a polar Sun-synchronous orbit. Launched in April 1970, this sensor continued to collect data until 1977. The instrument was a double monochromator which sampled 12 discrete wavelengths bands (Heath *et al.* 1973).

Residual Nimbus 4 b.u.v. hardware was later flown on the Atmospheric Explorer-5 spacecraft in November 1975, with an orbital inclination of 20°. Tropical data were collected and local time ozone changes were determined.

The most recent b.u.v. experiment is the Nimbus 7 s.b.u.v./t.o.m.s., launched in October 1978, in a Sun-synchronous polar orbit. This dual instrument continues the Nimbus 4 nadir ozone data set with the solar backscatter u.v. sensor (s.b.u.v.) (an improved version of the Nimbus 4 sensor) and adds spatial scanning of total ozone with the total ozone mapping spectrometer (t.o.m.s.).

Infrared emission methods for remote measurement of ozone are of two types (see table 1). Ozone profiles are determined from measurements of the altitude variation of the atmospheric radiance by viewing the Earth's limb (the limb radiance technique) and total ozone can be inferred from measurements of the atmospheric radiance in the nadir direction. The nadir sounding method was first tested on the Nimbus 3 satellite with an infrared interferometer spectrometer (i.r.i.s.) (Hanel *et al.* 1970). This was followed by a repeat flight on Nimbus 4. The retrieval methods and results were reported by Prabhakara *et al.* (1976).

The limb emission method is a more recent development, which has the advantages of good vertical resolution and simultaneous information about air temperature. Retrieval methods have been developed by Gille & House (1971) and the system has flown on the Nimbus 6 and 7 satellites.

Occultation techniques and results

Limb absorption or occultation measurements have been made with either the Sun or stars as the light source. Ozone profiles have been determined between about 40 and 100 km, thus giving the technique the distinction of working to higher altitudes than any other current sounding technique. Atmospheric attenutation is measured and subsequently interpreted in terms of Rayleigh scattering, atmospheric refraction and ozone slant path optical depth. Stellar occultations can, in principal, be obtained at any local time, although only night observations have been made with existing instrumentation. They are capable of high vertical resolution but are limited by the availability of suitable stars.

Solar occultation measurements are possible from a spacecraft sensor within an hour of sunrise or sunset at the ground twice during each orbit. The geographic coverage is rather severely limited by the orbit geometry. The vertical resolution is about 15 km unless the field of view is limited with a pointed telescope. The useful altitude range is determined by the number of wavelength channels, and the accuracy does not depend on absolute system calibration, but only on an electronic linearity and noise, and errors in effective ozone absorption coefficients.

The published results from occultation experiments are listed with references in table 1 and, since the quantity of soundings is small, the principal results are shown in figure 1 where the profiles are identified by the satellite name.

The K–M mid-latitude ozone model, (Krueger & Minzner 1976) which, for altitudes above 60 km, is derived primarily from twilight rocket flights, is also shown for comparison. The satellite profiles apply to a variety of latitudes, seasons and phases of the solar cycle and, as noted above, were obtained under near twilight or night conditions.

Night-time profiles from the orbiting astronomical observatories, OAO-2 and -3 show strong secondary maxima centered near 85 km and 95 km respectively. The OAO-3 results have been somewhat enigmatic because, at least below 80 km, the ozone densities are more than a factor of 2 greater than the results from the twilight experiments. If true, they would imply a night-time doubling in ozone density even at 50 km. The OAO-2 results exhibit the opposite effect below 75 km, namely a night-time ozone decrease relative to twilight measurements. The band of observed profiles from OAO-2 lies about a factor of 4 less than the K–M mid-latitude model at 50 km and appears to diverge at lower altitudes. The differences between the two OAO experiment results, and between them and the twilight results, are much greater than has been predicted for diurnal variations in the stratosphere and mesosphere.

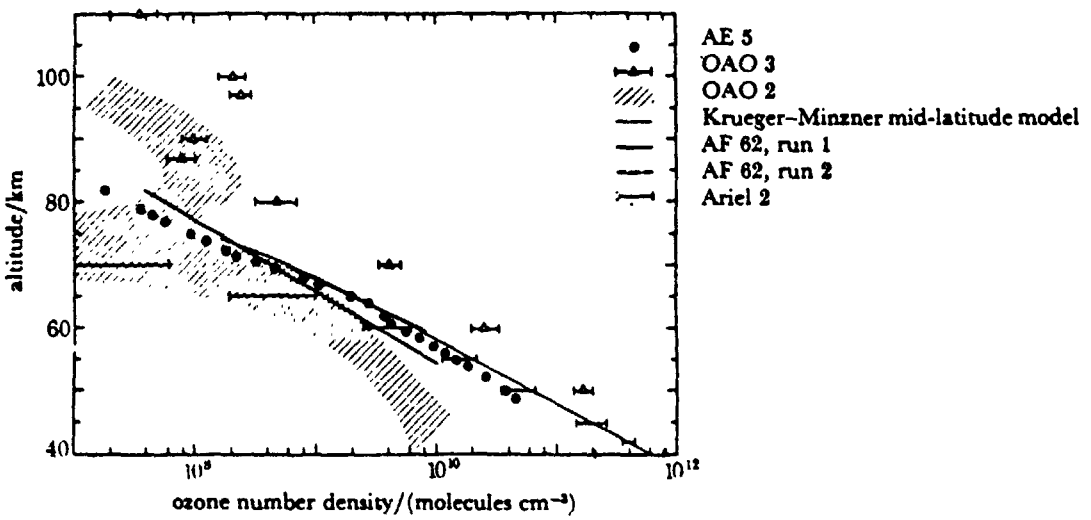


FIGURE 1. A comparison of ozone density distributions obtained with solar and stellar occultation methods.

The AE-5, AF 62, and Ariel 2 experiments are of the solar occultation type and, except for Ariel 2 at upper levels, the results tend to agree with the K–M mid-latitude model. The Ariel 2 data diverge toward lower values at higher altitudes and are consistent with the OAO-2 data at 65 and 70 km. The AE-5 results are 20–50 % less than the K-M mid-latitude model below 65 km. At those levels the K–M model becomes progressively more weighted toward daytime values. The AF 62 experiment determined only a two parameter profile, but two separate analysis techniques yielded different profiles (run 1 and run 2 in figure 1). The ozone densities from the AF 62 run 2 are quite close to AE-5 and K–M model results.

In summary, the solar occultation experiments differ significantly between themselves and from the rocket-derived K–M mid-latitude model. The ozone distribution shows a monotonic decrease with altitude with a scale height between 4 and 5 km for altitudes between 50 and 80 km. The ozone densities decrease from perhaps 5×10^{10} cm^{-3} at 50 km to about 1×10^{9} cm^{-3} at 75 km. The differences from the K–M model at 50 km may be due to latitude gradients, since most of the occultation results come from near the equator.

The stellar occultation results, taken at night, would require an extraordinarily variable

mesosphere to account for the 10- to 100-fold differences in their profiles. These profiles are particularly hard to accept near 50 km, where a relative wealth of other data exists.

Clearly, the occultation techniques are still in a developmental stage and mesospheric ozone profiles and fields have not been adequately determined.

The backscatter ultraviolet (b.u.v.) method

The b.u.v. method has proved to be a useful technique for observing global characteristics of the high altitude ozone field. The technique is successful owing to the intense ozone absorption continuum in the middle ultraviolet, the apparent lack of contamination of this band by interfering species and the demonstrated lifetime of the satellite hardware. The principal shortcoming is the inability to resolve-structure in the ozone profile less than about an atmospheric scale height. This, incidentally, is not a severe drawback due to the general 'well behaved' nature of the ozone distribution in the upper stratosphere. However, no information is returned about the ozone distribution below the ozone maximum, although total ozone can be determined.

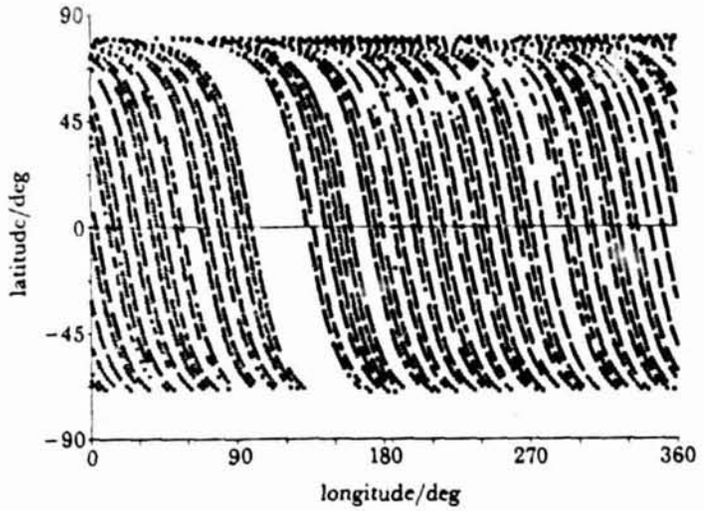


Figure 2. Data coverage for a typical week of b.u.v. data. Each square represents a b.u.v. sounding.

The inference of total ozone amounts and its vertical distribution is possible by solution of an integral equation. The fundamental observed parameters are the incident u.v. solar spectral irradiance and the atmospheric spectral radiance in the middle u.v. (250–340 nm). In addition, the surface reflectance must be determined by measurements at a wavelength that is free of ozone absorption. The wavelength dependence of the ozone absorption coefficient and the exponential decrease of air density with height leads to a series of contribution functions peaking at different levels corresponding to different wavelengths between 250 and 300 nm. The ozone profile can be inferred by using this altitude separation. At wavelengths greater than 300 nm, the contribution functions peak below the stratosphere and a total ozone measurement is possible.

Design of a spectrometer to measure the atmospheric radiance is difficult because the Earth's albedo changes drastically with wavelength between 250 and 340 nm, owing to ozone absorption. Furthermore, the solar flux is a factor of 10 less at 250 nm that it is at 340 nm. As a result, in a spectral scan of the earth radiance, the flux changes by 10^4. Further, observational

complications come from polarization of the backscattered light. To meet these stringent observational requirements, the Nimbus b.u.v. instruments are double (tandem) Ebert–Fastie monochrometers with collinear narrow-band filter photometers. The instrument views the terrestrial atmosphere at the nadir. At the northern terminator, a diffuser plate is deployed to compare the solar spectral irradiance to the atmospheric radiance.

The Nimbus 4 Backscatter Ultraviolet (b.u.v.) experiment was launched in April of 1970. The spacecraft is in a near polar Sun synchronous orbit and data are available over almost all of the sunlit portion of the earth with the observations taken near local noon. Coverage for a typical week is shown in figure 2 where each 200 km by 200 km sounding is represented by ×. There are an average of 800 data points per day for the first year of operations. Increasingly severe spacecraft power problems starting in July 1972 have caused a substantial reduction in operating time for the instrument although data are still occasionally collected from it.

Vertical profile results from b.u.v. observations

The first experiment in the development of the backscatter u.v. methods used a single channel photometer. Rawcliffe & Elliott (1966) inferred from the radiance data that seasonal differences were present in upper stratospheric ozone. The first findings from an ozone profile experiment using an ultraviolet spectrometer were that the upper stratospheric ozone scale height was constant at sub-polar latitudes and that detailed horizontal structure existed near 30 mbar† (Iozenas *et al.* 1969). These results indicated that photochemistry prevailed over dynamics at the upper levels. In succeeding experiments with spectrometers, it was suggested that instrument calibration uncertainties or possible aerosol scattering might present real problems for accurate estimation of ozone profiles (Elliott 1971). However, Anderson *et al.* (1969) found that profiles could be recovered after calibration with rocket ozone data and that latitudinal and seasonal ozone variations could nevertheless be determined (London *et al.* 1977). In the two data periods (Sept.–Nov. 1967 and Jan.–Feb. 1968) that have been analysed from the OGO-4 experiments, the average ozone fields at 5 mbar were found to have small geographic variations. A broad maximum appears at mid-latitudes and, in the tropics, some longitudinal asymmetry is present with higher ozone mixing ratios over the Atlantic and Africa. These geographical variations diminish with altitude in agreement with the earlier observations. The mixing ratios at higher latitudes were found to increase from autumn to winter.

Initial analysis of the Nimbus 4 ozone profile data (Krueger *et al.* 1973) revealed that the highest mixing ratios were near 10 mbar at lower latitudes. During the equinoxes, the mixing ratios decrease rather uniformly toward the poles. However, during the autumn and winter secondary maxima form near the 2 mbar level at Southern Hemisphere mid-latitudes and at Northern Hemisphere high latitudes. Perhaps the simplest explanation of these upper stratospheric features lies in the temperature dependence of the ozone chemistry. The southern maximum, in particular, was found to vary in amplitude at different longitudes. Barnett *et al.* (1975) used these b.u.v. results, together with s.c.r. temperature data, to estimate the temperature coefficient for steady state ozone mixing ratios. They concluded that the exponential coefficient was of the order of 900–1100 K, a value consistent with oxygen and hydroxyl chemistry near 1 mbar.

The Nimbus 4 b.u.v. was also taking data during the major solar proton event of August 1972. Ozone decreases of 20 % were found at 2 mbar in the upper stratosphere at geomagnetic

† 1 mbar = 10^2 Pa.

latitudes irradiated by protons (Heath *et al.* 1977). This was the first direct evidence for catalysis of ozone by odd nitrogen compounds in the stratosphere.

Methods for improvement of retrievals of high level ozone profiles from reprocessed b.u.v. Earth albedo data are currently under study. This new data set incorporates corrections for instrument sensitivity changes and spacecraft attitude errors. The aerosol effects found earlier by Elliott (1971) do not appear in Nimbus 4 b.u.v. data, although anomalous radiances appear to exist at 255 nm. Recently, the global behaviour of 1 mbar ozone from April 1970 to May 1972 has been studied at Goddard Space Flight Center by using an interim inversion product that is valid for the upper stratosphere – Mateer's formulation of the Laplace transform method (Twomey 1961). The variation of the zonal mean of the 1 mbar ozone mixing ratio at 60 N for the period Nov. 1970–Nov. 1971 is shown in figure 3. The 2 mbar air temperatures for the same latitude band and time period derived from s.c.r. observations (Barnett 1974) are shown in the same figure. Both b.u.v. and s.c.r. observe the radiation from rather broad height regions so the excellent inverse relationship of temperature and ozone is quite pleasing.

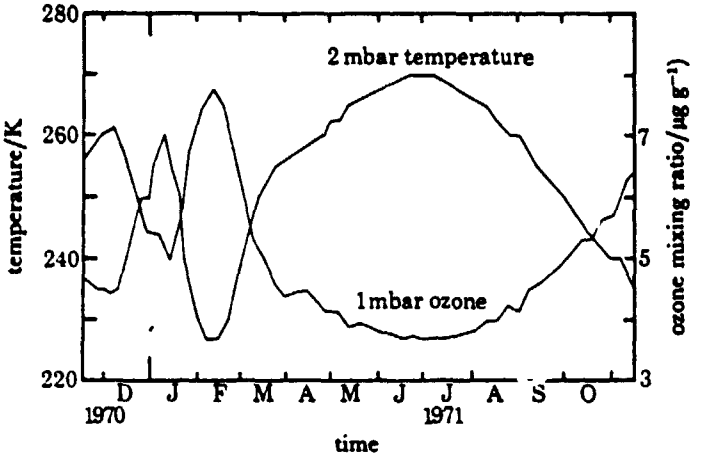


FIGURE 3. Temporal variation of the 2 mbar temperature from s.c.r. and the 1 mbar ozone mass mixing ratio in a latitude band centred at 60° N from 1 year of Nimbus 4 data.

The overall behaviour observed is a summer temperature maximum (270 K) and ozone minimum (3.71 μg/g) and a winter temperature minimum (227 K) and ozone maximum (7.62 μg/g). The large sudden stratospheric warming from mid-December to mid-January is mirrored exactly as a decrease in 1 mbar ozone. Even a small scale feature at the beginning of September is observed in both temperature and ozone. Temperature (T) and the ozone mixing ratio (r_3) are found to be related by the equation

$$r_3 = 0.072 \exp(1062/T)$$

with a correlation coefficient of 0.991 for the fit. This relation holds for the entire year except during the stratospheric warming, when the pre-exponential coefficient 0.072 μg/g increases to 0.082 μg/g. The temperature coefficient of 1062 K compares reasonably well with the value of 928 K reported by Barnett *et al.* (1975) at 50 S, but the correlation coefficient they report of 0.817 is substantially lower. This is due in part to our using much better calibrated ozone data and in part to using zonally averaged data. The increase in the pre-exponential coefficient during the stratospheric warming period is evidence for modification of the chemistry and dynamics from that normally present. This could happen if significant vertical motions arise.

Data from the AE-5 b.u.v. have been partially analysed. Principal findings to date are:

(1) the tropical profiles are consistent with current photochemical model results below 2 mbar but not higher altitudes (Frederick *et al.* 1978).

(2) the transition from dynamic to photochemical control of the ozone takes place between 5 and 10 mbar in the tropics (Frederick *et al.* 1977*b*) and

(3) negligible spatial structure exists in the integral ozone above 1 mbar at tropical latitudes when measured with 25 km horizontal resolution (Guenther & Dasgupta 1979).

TABLE 2. B.U.V./DOBSON OVERALL COMPARISON STATISTICS CHART

Year 1 (April 1970–April 1971)	number of coincidences	$\bar{\Omega}_N$ (m atm-cm)	$\bar{\Omega}_D$ (m atm-cm)	bias $(\bar{\Omega}_N-\bar{\Omega}_D)$	s.d.
Dobson 00 code, 2° sep, all R;	2024	315.9	326.4	−10.6	18.44
1° sep, $R \leqslant 0.2$	328	304.9	316.6	−11.8	16.28
all codes	3860	325.6	336.4	−10.9	21.29
all codes exc. 00 code	1841	336.4	347.7	−11.3	24.02
M-83 Network†	1270	349.7	358.8	−9.1	59.50
Year 2 (May 1971–April 1972)					
Dobson 00 code, 2° sep, all R;	1944	304.0	317.9	−13.9	18.83
1° sep, $R \leqslant 0.2$	301	292.0	304.4	−12.4	16.26
all codes	3392	313.9	329.2	−15.4	20.57
all codes exc. 00 code	1487	327.3	344.2	−17.0	23.65
M-83 network†	1117	342.3	368.9	−26.6	51.51

† Non-Dobson Russian filter photometer.

Total ozone from b.u.v. observations

The first application of the Dave & Mateer (1967) method for total ozone estimation from u.v. earth albedo measurements was on the Nimbus 4 b.u.v. Samples of the data were processed to establish the feasibility of the technique (Mateer *et al.* 1971) and to indicate the nature of time variations in the global ozone field (Krueger 1974). In 1976, as a result of the increasing importance of accurate monitoring of global ozone, a special effort was organized to process all the available data with the best possible accuracy. This effort included development of improved processing algorithms, compensation for instrument changes with ageing, careful review of the spacecraft ephemeris and attitude data and quality control of the output data. All the total ozone data have been reprocessed and archived in the National Space Sciences Data Center. Specific information as to how to acquire the processed Nimbus 4 total ozone data may be obtained by writing to World Data Center for Rockets and Satellites, Code 601, Goddard Space Flight Center, Greenbelt, Md. 20771, United States of America. There has been an extensive effort to assess the quality of the resulting data set. The primary comparison has been with the ozone measurements obtained from *Ozone data for the world* as provided by the Atmospheric Environment Services of Canada. All available spacecraft and ground data collocations have been identified and a variety of statistics computed. Results from several cases for the first 2 years are shown in table 2. All possible cases meeting the collocation criteria are included in each line with no preselection of either the ground or spacecraft data. The first line for each year is for AD-direct Sun wavelength pair (00 code) observations taken on the same day, with the centre of the spacecraft field-of-view within 2° of the ground station. A scatter plot of this set of data is shown in figure 4. The second line in the table 2 categories is a subset of the first,

with a colocation limit of 1° and only cases where the spacecraft scene is cloud free are included. The third line includes all types of Dobson data and the fourth line is a subset of the third in which all except the 00 code data are included. A similar comparison of b.u.v. data and M-83 data is shown in the last line for each year. There are four major types of errors that contribute to the overall variance between b.u.v. and ground data shown in this table: random errors in the b.u.v. data; random errors in the ground data; noise introduced by imperfect colocation in time, space and area sampled; and noise resulting from the apparent biases of individual ground stations as shown in the Dobson intercomparisons performed at Belsk and Boulder. It is not possible to establish how much of the bias between a specific ground station and the b.u.v. data originates in the ground station; however, is can be shown that roughly half the overall variance can be removed by nulling these biases out on a station by station basis. When reasonable estimates are made for each of the types of errors mentioned above, it appears that the b.u.v. data is similar in accuracy to a well run Dobson station.

Seasonal variations of total ozone have now been recomputed from the current b.u.v. data sets (Hilsenrath *et al.* 1978). A time–latitude cross section of zonally averaged total ozone was generated from average daily means for 10° wide latitude zones. In figure 5, the total ozone has been contoured with 20 m atm-cm increments for the 26-month period from April 1970

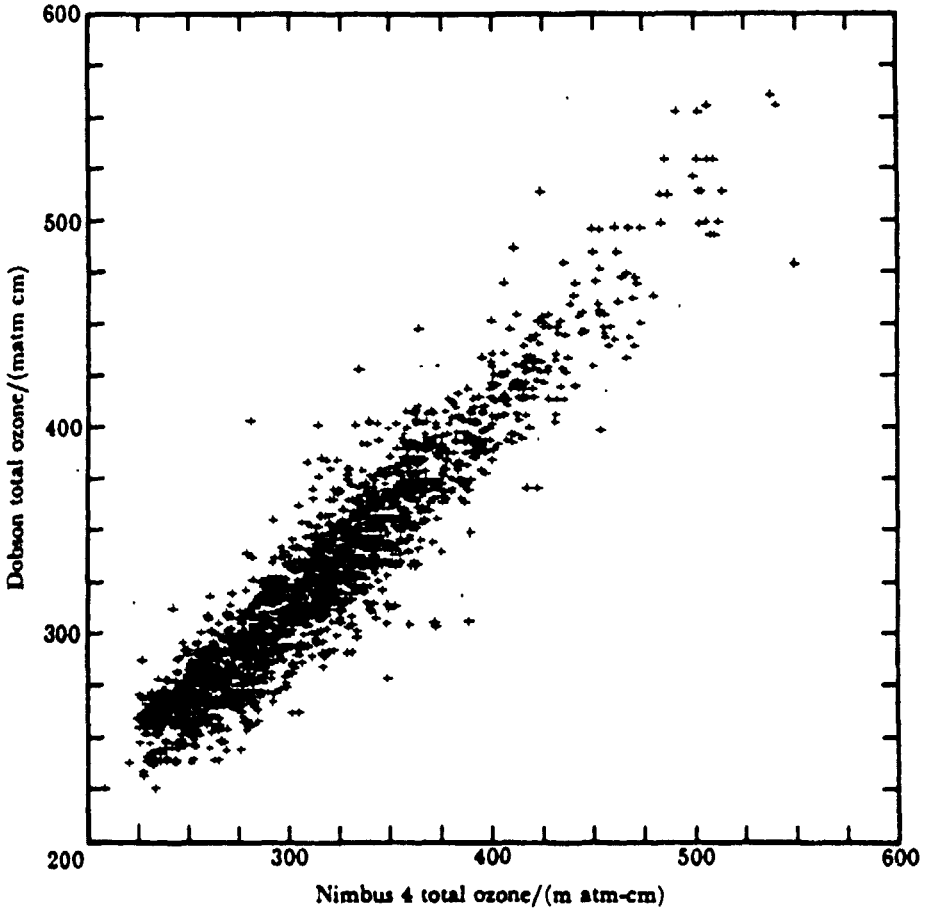


FIGURE 4. Comparison of b.u.v. total ozone estimations with coincident Dobson spectrophotometer total ozone measurements during the period 10 April 1970 to 30 April 1971. Only ADDS Dobson observations are included.

to May 1972 (contours are missing in the polar night because there are no observations during that period). The seasonal march of total ozone in the two hemispheres is clearly seen. A comparison with analyses of Dobson data (see, for example, London *et al.* 1976) reveals close correspondence with the major features in the ozone climatology. The b.u.v. results point up the distinctive differences in the seasonal trends for each hemisphere. These differences are as follows:

(1) In the Northern Hemisphere the spring maximum occurs nearly simultaneously at mid and high latitudes. The maximum value occurs near the pole, with total ozone amounts of about 500 m atm-cm in 1970 and decreasing to 460 and 440 m atm-cm in the spring of 1971 and 1972.

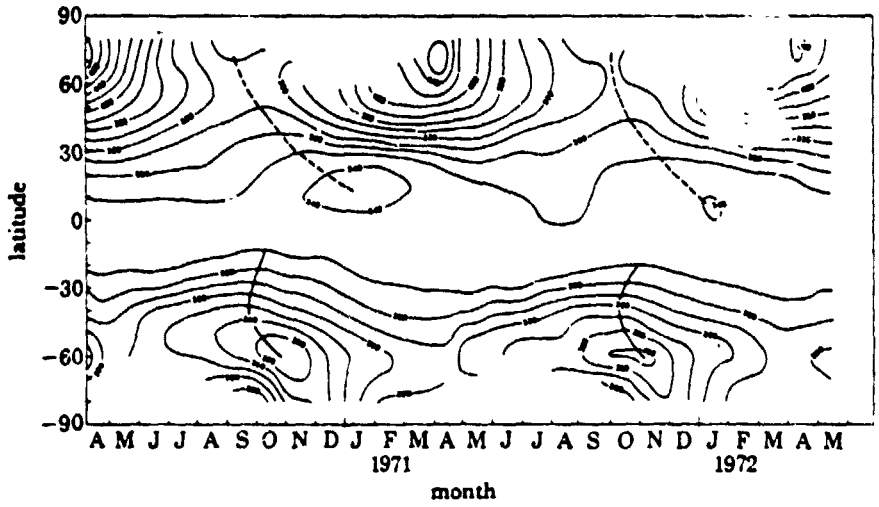


FIGURE 5. Time–latitude cross section of zonally averaged total ozone obtained from the Nimbus 4 b.u.v. from April 1970 to May 1972. The values are given in m atm-cm of ozone

(2) In the Southern Hemisphere the spring build-up begins at 50 S in September with ozone values substantially lower than those in the Northern Hemisphere spring. The maximum values occur about 1 month later at polar latitudes. The asymmetry in the winter build-up in the two hemispheres is indicative of differences in the circulation features in the two hemispheres.

(3) In the tropical regions, the average total ozone amount is about 250 m atm-cm and a small seasonal variation is present, showing more rapid changes than in the long term Dobson averages. The ozone minimum, which is centred below the equator in April in both years moves northward after August. The lowest ozone amounts, which are found in the northern tropics, appear to follow after Southern Hemisphere spring maximum and are coincident with the Northern Hemisphere winter build-up.

A harmonic analysis has been made of the zonal mean data for the 2-year period (Hilsenrath *et al.* 1978). The results of this analysis are the following:

(1) More than 90 % of the variance of the zonal mean values at most latitudes is found in the two harmonics that represent the annual and semi-annual waves. The search for periodicities shorter than 1 month found no amplitudes greater than 5 m atm-cm at mid to high latitudes and 2 m atm-cm at low latitudes.

(2) At mid to high latitudes the annual wave dominates in both hemispheres. In the Northern Hemisphere the amplitude of the annual wave increases with latitude while in the Southern

Hemisphere the peak amplitude occurs at 50 S in the first year and at 60 S in the second year. The amplitude of the annual wave is a minimum at the equator in the first year and at 10 S in the second year. This shift in the minimum of the annual wave could be a result of the year to year change in the strength and location of the Hadley Cell circulation in the hemisphere.

(3) The semi-annual wave was observed at the higher latitudes in both hemispheres for the 2 years and in the tropics in the first year. This pattern is very similar to the analysis by London *et al.* (1976) of the Dobson data base. However, compared with the annual wave, the fractional variance in the semiannual wave is significant only in the tropics and at high southern latitudes. The tropical amplitude is small and probably is only an expression of the annual waves in the two hemispheres and, indeed, is in phase with these waves. The amplitude of the polar semiannual wave is about 20 m atm-cm. This is comparable to the amplitude of the annual wave at 80 S but only a quarter of that amplitude at 80 N.

Infrared remote sensing of total ozone

The Nimbus 4 satellite carried an infrared interferometer spectrometer (i.r.i.s.), which provided measurements of the thermal emission spectrum of the Earth and atmosphere between about 400 and 1600 cm^{-1} with an apodized spectral resolution of 2.8 cm^{-1} (Hanel *et al.* 1971). The field of view of the instrument projected on the surface of the Earth was a circle with a diameter of approximately 95 km and all data were acquired near either local noon or local midnight between 80 N and 80 S. Measurements were essentially continuous from April to December 1970 and were available for a few days in January 1971.

The spectral measurements of i.r.i.s. in the 15 μm CO_2 band were inverted to derive the vertical distribution of temperature in the atmosphere (Conrath 1972). Using these temperature data, together with i.r.i.s. spectral measurements in the 9.6 μm ozone band, Prabhakara *et al.* (1976) deduced the global distribution of total ozone for about 10 months.

Though the i.r.i.s. was not a dedicated experiment for ozone soundings, the 9.6 μm band basically yields one parameter about ozone. This parameter depends both on the total ozone and its vertical distribution. Therefore, additional information must be provided to obtain total ozone explicitly. This additional information is provided by an empirical relation, based on balloon observations, which implies that increases in total ozone occur mainly because of increases in ozone concentration in the layers of the atmosphere below the ozone maximum. This empiricism and the crude vertical resolution, approximately 1 scale height, of the remotely sensed temperature profile in the lower stratosphere lead to a basic limitation on the infrared ozone remote sensing technique described by Prabhakara *et al.* (1976).

Since the time of the i.r.i.s flights, several temperature sounding instruments have included a channel to sense the 9.6 μm ozone band. Beginning in March 1977, the multichannel filter radiometer instruments on four of the Defense Meteorological Satellite Program series of satellites has included such a channel. Lovill *et al.* (1978) have reported preliminary results. The most recent 9.6 μm ozone sensor is part of the high resolution infrared radiometer (h.i.r.s.) sensor on the Tiros N Satellite, launched in November 1978.

Despite the limitations of the nadir infrared method, good coverage of the satellite data over the globe permit us to get some gross information on the geographic and seasonal variation of total ozone. The sensing capability at night and in the polar night could be very useful.

In figure 6, a map of the global distribution of ozone derived from i.r.i.s. for the month of December 1970 is shown to highlight the ozone information over the northern polar latitudes

in winter. Longitudinal changes of about 100 m atm-cm can be seen between the maximum values east of Siberia and the minimum values in the North Atlantic. In the tropics the lowest values appear in the Indonesian sector, perhaps in association with the Siberian ozone high. The tropical ozone tends to have a wavenumber 1 characteristic throughout the year with the maximum over the Atlantic as shown in the December chart. The maximum amplitude of this wave is in July and August.

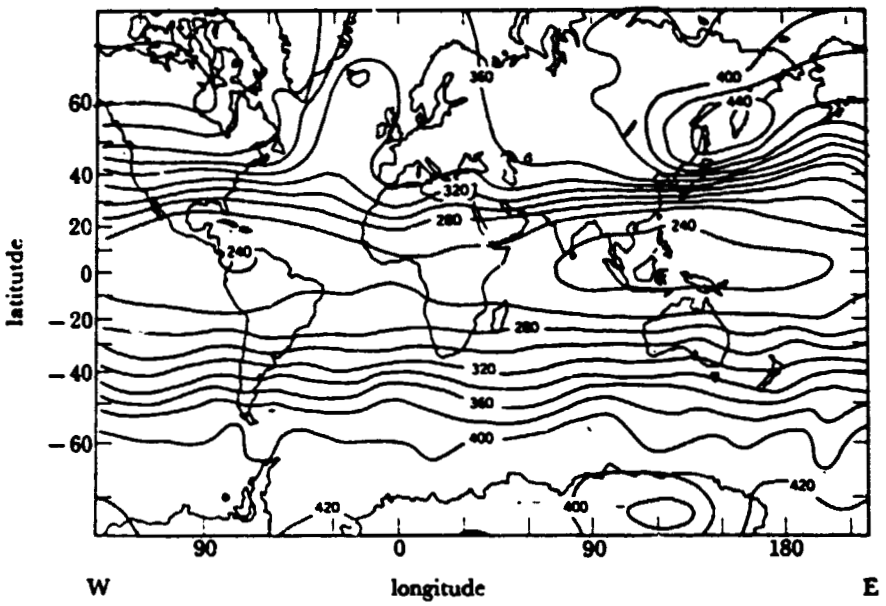


FIGURE 6. Mean global distribution of total ozone for December 1970 obtained from the Nimbus 4 i.r.i.s. instrument. The values are given in m atm-cm of ozone.

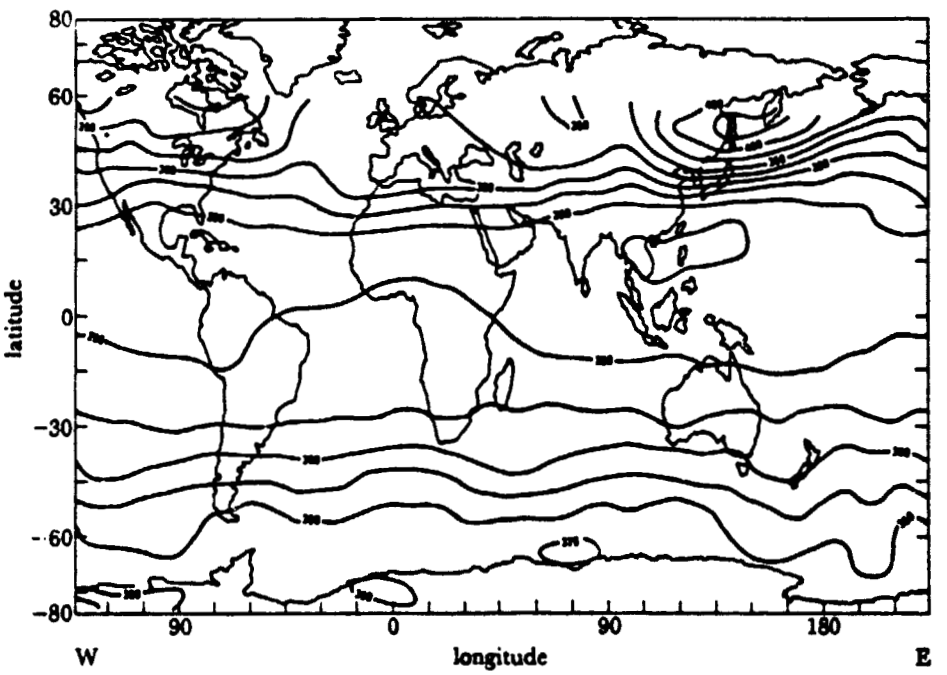


FIGURE 7. Mean global distribution of total ozone for December 1970 obtained from the Nimbus 4 b.u.v. instrument. The values are given in m atm-cm ozone. This chart may be compared with figure 6.

In the Southern Hemisphere, the December (summer) ozone field is fairly uniform at all latitudes except for the Antarctic regions. The i.r.i.s. December map exhibits a 40 m atm-cm zonal variation with the minimum located at the same longitude as the tropical minimum.

Comparative results from b.u.v. and i.r.i.s.

A comparison of the ozone results from i.r.i.s. and b.u.v. is of interest. Figure 7 is a December map of total ozone derived from the Nimbus 4 b.u.v. instrument. This can be compared directly with the map derived from i.r.i.s. (figure 6). In general the observed ozone features from the two instruments are the same. Taking into account the different contour intervals in the two maps, however, some differences in the absolute values can be found. This is expected from the study performed by Prabhakara *et al.* (1976), who demonstrate a reduction in the seasonal amplitude of the i.r.i.s. total ozone amounts. The two satellite measurements compare well in the tropics. The absolute amounts are very close and the minima occurring over Central America and Indonesia are observed by both instruments. In the Southern Hemisphere both measurements show a uniform latitude gradient from low to high latitudes; however, the b.u.v. values are about 15 % lower than the i.r.i.s. values. In the Northern Hemisphere the ozone high east of Siberia and the features in the North Atlantic are detected by both instruments. The absolute values agree here a little better than in the Southern Hemisphere.

Although the two instruments were on the same satellite, the data bases for these maps are not identical. B.u.v. observations are possible only in daytime, while i.r.i.s. soundings are taken day and night and, in addition, during the polar night. Furthermore, the instrument duty cycles were not identical and the data recovery varied between i.r.i.s. and b.u.v. However, it is encouraging that the features are reproduced since the measurement techniques are vastly different.

Conclusions

During the 1960s and into the 1970s, several methods for remote measurement of ozone were devised and tested. The most promising techniques for stratospheric soundings in the 1980s seem to be the backscatter u.v. system for ozone profile and total ozone monitoring and the limb infrared emission system for investigations requiring good vertical resolution of the ozone profile and complementary information on air temperatures and trace constituents. For mesospheric ozone measurements, the occultation methods have shown promise, although the unreasonable differences existing between experimental results must be resolved.

At present, several data sets have been collected and climatological descriptions are being assembled. The interpretation of these results in terms of atmospheric processes is only beginning. Already, photochemical modelling has been guided and theories tested by the satellite data. The ultimate utility will perhaps be found after the chemistry is well understood. At that point the global ozone data can be used for tracing air motions as a complementary source of information for dynamic models of the atmosphere.

References (Krueger *et al.*)

Anderson, G. P., Barth, C. A., Cayla, F. & London, J. 1969 *Ann. Geophys.* **25**, 239–243.

Barnett, J. J. 1974 *Qt. Jl R. met. Soc* **100**, 505–530.

Barnett, J. J., Houghton, J. T. & Pyle, J. A. 1975 *Jl R. met. Soc* **101**, 245–257.

Conrath, B. J. 1972 *J. atmos. Sci.* **29**, 1262–1271.

Dave, J. V. & Mateer, C. L. 1967 *J. atmos. Sci* **24**, 414–427.

Elliott, D. D., Clark, M. A. & Hudson, R. D. 1967 *Aerospace Techn. Report no.* TR-0158 (3260-10), 2 September 1967.

Elliott, D. D. 1971 *Space Research* XI, pp. 857–861. Berlin: Akademie-Verlag.

Frederick, J. E., Hays, P. B., Guenther, B. W. & Heath, D. F. 1977*a* *J. atmos. Sci.* **34**, 1987–1994.

Frederick, J. E., Guenther, B. W. & Heath, D. F. 1977*b* *Beitr. Phys. Atmos.* **50**, 496–507.

Frederick, J. E., Guenther, B., Hays, P. B. & Heath, D. F. 1978 *J. geophys. Res.* **83**, 953–958.

Gille, J. C. & House, F. B. 1971 *J. atmos. Sci.* **28**, 1427–1442.

Guenther, B. & Dasgupta, R. 1979 In preparation.

Guenther, B., Dasgupta, R. & Heath, D. F. 1977 *Geophys. Res. Lett.* **4**, 434–437.

Hanel, R. A., Schlachman, B., Clark, F. D., Prokesh, C. H., Taylor, J. B., Wilson, W. M. & Chaney, L. 1970 *Appl. Opt.* **9**, 1767–1774.

Hanel, R. A., Shlachman, B., Rogers, D. & Vanouse, D. 1971 *Appl. Opt.* **10**, 1376–1382.

Hays, P. B. & Roble, R. G. 1973 *Planet. Space Sci.* **21**, 273–279.

Heath, D. F., Mateer, C. L. & Krueger, A. J. 1973 *Pure appl. Geophys.* **106–108**, 1239–1253.

Heath, D. F., Krueger, A. J., Roeder, H. A. & Henderson, B. D. 1975 *Opt. Engng* **14**, 323–331.

Heath, D. F., Krueger, A. J. & Crutzen, P. J. 1977 *Science N.Y.* **197**, 323–331.

Hilsenrath, E., Heath, D. F. & Schlesinger, B. M. 1979 *J. geophys. Res.* **84**. (In the press.)

Iozenas, V. A., Krasnopol'skiy, V. A., Kuznetsov, A. P. & Lebedinsky, A. I. 1969 *Atmos. oceanic Phys.* **5**, 149–159.

Krueger, A. J. 1974 *Proc. Int. Conf. on Structure, Composition and General Circulation of the Upper and Lower Atmosphere and Possible Anthropogenic Perturbations*, vol. 1, pp. 467–477.

Krueger, A. J., Heath, D. F. & Mateer, C. L. 1973 *Pure appl. Geophys.* **106–108**, 1254–1263.

Krueger, A J. & Minzer, R. A. 1976 *J. geophys. Res.* **81**, 4477–4481.

London, J., Bojkov, R. D., Oltmans, S. & Kelley, J. I. 1976 *N.C.A.R. Tech. Note* TN/113+STR. National Center for Atmospheric Research, Boulder, Colorado.

London, J., Frederick, J. E. & Anderson, G. P. 1977 *J. geophys. Res.* **82**, 2543–2556.

Lovill, J. E., Sullivan, T. J., Weichel, R. L., Ellis, J. S., Huebel, J. G., Korver, J. A., Werdhaas, P. P. & Phelps, F. A. 1978 *Lawrence Livermore Laboratory* UCRL-52473, May 25.

Mateer, C. L. 1972 In *Mathematics of profile inversion* (ed. L. Colin), pp. 2–25. N.A.S.A.TMX-62,150.

Mateer, C. L., Heath, D. F. & Krueger, A. J. 1971 *J. atmos. Sci.* **28**, 1307–1311.

Miller, D. E. & Steward, K. H. 1965 *Proc. R. Soc. Lond.* A288, 540–544.

Nimbus Project 1978 *The Nimbus 7 Users Guide*, NASA, Goddard Space Flight Center, Greenbelt, Md.

Prabhakara, C., Rodgers, E. B., Conrath, B. J., Hanel, R. A. & Kunde, V. G. 1976 *J. geophys. Res.* **81**, 6391–6399.

Rawcliffe, R. D., Meloy, G. E., Friedman, R. M. & Rogers, E. H. 1963 *J. geophys. Res.* **68**, 6425–6429.

Rawcliffe, R. D. & Elliott, D. D. 1966 *J. geophys. Res.* **71**, 5077–5089.

Riegler, G. R., Drake, J. F., Liu, S. C. & Cicerone, R. J. 1976 *J. geophys. Res.* **81**, 4997–5001.

Singer, S. F. & Wentworth, R. C. 1957 *J. geophys. Res.* **62**, 229–308.

Twomey, S. 1961 *J. geophys. Res.* **66**, 2153–2162.

Venkateswaran, S. V., Moore, J. G. & Krueger, A. J. 1961 *J. geophys. Res.* **66**, 1751–1771.

Reprinted from APPLIED OPTICS, Vol. 18, page 3421, October 15, 1979

Tropospheric UV flux calculations and photolysis rates for use with zonally and diurnally averaged models

Gregory L. Matloff and Richard W. Stewart

A computationally fast and efficient method for calculating solar UV fluxes in the troposphere is developed. Calculated fluxes compare favorably with more rigorous multiple scattering results, and qualitatively correct behavior of the direct and diffuse flux components is maintained. In view of the relatively large uncertainties in other aspects of tropospheric photochemical models, the use of approximate UV flux calculations appears warranted. The technique is applied to the calculation of photolysis rates for twelve significant photolytic reactions of nine species.

I. Introduction

Many problems of current interest in climatology and environmental quality require an understanding of the trace species composition of the global-scale troposphere.[1] This composition is determined largely by the photolysis of ozone, nitrogen dioxide, and several other species by UV radiation in the 290–400-nm region. The calculation of the associated photolysis rates is an essential element in photochemical model descriptions of the troposphere.

The many uncertainties in chemical, radiative, and physical parameters which enter into model calculations require that a range of sensitivity studies be carried out to determine model response to reasonable parameter variations. This requirement can only be met if no single detail of the computational scheme is severely rate limiting. The rigorous calculation of tropospheric fluxes and photolysis rates can potentially limit the range of model sensitivity studies, since it involves multiple scattering calculations that can consume substantial amounts of computer time. To avoid such difficulties and because the effect of neglecting detailed radiative transfer considerations on computed photolysis rates was unknown, earlier photochemical models of the troposphere and stratosphere tended to treat molecular absorption as the only UV flux attenuating mechanism. However, recent multiple scattering calculations[2–4] have shown that other physical interactions of the solar flux with the atmosphere and ground can significantly influence the values obtained for the ambient UV flux. It is, therefore, desirable to use flux and photolysis rate calculations that give reasonably accurate results when compared with more rigorous techniques, avoid the time consuming complexities of such techniques, and are readily incorporated into existing global tropospheric models. The development of such a calculation is the objective of this paper.

Gregory Matloff is with Systems & Applied Science Corporation Riverdale, Maryland 20840, and Richard Stewart is with NASA Goddard Space Flight Center, Goddard Laboratory for Atmospheric Sciences, Greenbelt, Maryland 20771.

Received 18 September 1978.

II. Computational Method

A. Fluxes

The solar UV flux at wavelength λ and altitude Z in the troposphere will be expressed as the sum of direct, scattered, and reflected components:

$$F_\lambda(Z) = F_{D\lambda}(Z) + F^{\uparrow}_\lambda(Z) + F^{\downarrow}_\lambda(Z) + F_{R\lambda}(Z), \quad (1)$$

where $F^{\uparrow}$ and $F^{\downarrow}$ indicate forwardscattered and backscattered fluxes, respectively. The flux and its components are also functions of latitude and season, but explicit dependence on these factors is suppressed in Eq. (1) to simplify the notation.

The direct flux is attenuated by molecular absorption and scattering and by aerosol scattering. In addition it is modified by factors expressing the mean cloud cover and the sunlit fraction of the day (diurnal average). The direct flux in our formulation is thus expressed as

$$F_{D\lambda}(Z) = F_0(\lambda)D(\psi,\delta)[1 - f(\psi,\delta)\gamma(Z)]T_{oz} \cdot T_a \cdot T_p, \quad (2)$$

where $F_0(\lambda)$ is the UV flux at the top of the atmosphere, $D(\psi,\delta)$ is the sunlit fraction of the day as a function of latitude ψ and solar declination δ, $f(\psi,\delta)$ is the mean fractional cloud cover, and $\gamma(Z) = 0$ or 1 depending on whether $Z > Z_c$ or $Z < Z_c$, where Z_c is the associated cloud height. T_{oz}, T_a, and T_p are the transmissions due to ozone absorption, Rayleigh scattering, and aerosol

scattering. For convenience in the following, we define a modified flux by

$$F_0(Z) = F_0(\lambda)D(\psi,\delta)[1 - f(\psi,\delta)\gamma(Z)]T_{oz}. \quad (3)$$

We approximate the forwardscattered flux by assuming that half of the flux scattered out of the direct beam by the atmosphere above Z is returned to the downward flux by further scattering events:

$$F^+(Z) = \tfrac{1}{2}F_0(Z)(1 - T_a \cdot T_p). \quad (4)$$

We are implicitly assuming, in Eq. (4) and the following expressions, that ozone absorption occurs above the troposphere and that scattering provides the major attenuating mechanism below the tropopause.

The backscattered flux is approximated by the analogous assumption that half the flux scattered out of the direct beam between altitude Z and the ground is returned by further scattering. There is an added complication in this case, since the modified flux reaching the ground may be less than the modified flux at Z due to the presence of clouds. At altitude $Z > Z_c$, we must thus consider backscattered radiation from above and below the cloud layer. If $F_0(Z_c+)$ and $F_0(Z_c-)$ denote the modified fluxes just above and just below the clouds, for altitudes $Z > Z_c$ we have

$$\begin{aligned} F^-(Z) = {} & \tfrac{1}{2}F_0(Z_c+)[1 - T_a(Z_c)T_p(Z_c)] \\ & - \tfrac{1}{2}F_0(Z)[1 - T_a(Z)T_p(Z)] + [\tfrac{1}{2}F_0(0) \\ & \times [1 - T_a(0)T_p(0)] - \tfrac{1}{2}F_0(Z_c-)[1 - T_a(Z_c) \\ & \times T_p(Z_c)]]T(Z,Z_c). \end{aligned} \quad (5a)$$

while below the clouds

$$\begin{aligned} F^-(Z) = {} & \tfrac{1}{2}F_0(0)[1 - T_a(0)T_p(0)] \\ & - \tfrac{1}{2}F_0(Z)[1 - T_a(Z)T_p(Z)]. \end{aligned} \quad (5b)$$

$T(Z,Z_c)$ in Eq. (5a) is the transmission between altitude Z and Z_c and is given by

$$T(Z,Z_c) = T_a(Z)T_p(Z)/T_a(Z_c)T_p(Z_c) \quad (6)$$

For the reflected flux we assume

$$\begin{aligned} F_R(Z) = {} & G(Z_c) \cdot A_c \cdot f(\psi,\delta) \cdot T(Z,Z_c) \\ & + G(0) \cdot A_G(\psi,\delta) \cdot T(Z,0), \end{aligned} \quad (7)$$

where A_c and $A_G(\psi,\delta)$ are the cloud and ground albedos, and $G(Z)$ is the net downward flux at Z:

$$G(Z) = F_D(Z) + F^+(Z) \quad (8)$$

$T(Z,0)$ is the transmission between Z and the ground given by an expression analogous to Eq. (6).

B. Flux Modification Factors

Shettle and Green[2] (hereafter SG) and Nack and Green[3] have published multiple scattering calculations of the solar UV flux, which includes effects of molecular absorption and scattering, aerosol scattering, cloud attenuation, and ground albedo. To facilitate comparison of our results with these more rigorous calculations, we will use the same parameterizations of physical quantities as SG[2] whenever possible. Some of these will be extended to include latitude and seasonal dependence.

For the incident solar flux at the top of the atmosphere, we use the same approximation as SG:

$$F_0(\lambda) = K[1 + (\lambda - \lambda_0)/d], \quad (9)$$

with $K = 0.552$ W/m^2 nm, $\lambda_0 = 300$ nm, and $d = 37$ nm.

The factor $1 - f(\psi,\delta)\gamma(Z)$ in Eq. (2) accounts for the blocking of solar radiation by clouds. The mean fractional cloud cover given in *CIAP Monographs 1* and 4[5,6] may be described by the following:

$$f(\psi,\delta) = \begin{cases} 0.55 - 0.2\sin 3.6\psi & 90°S > \psi > 30°S & \text{spring} \\ 0.47 + 0.07\sin 9.0\psi & 30°S \geq \psi > 0° & \text{summer} \\ 0.6 - 0.2\sin 4.5\psi & 90°N > \psi \geq 0° & \\ 0.51 - 0.11\sin 4.5\psi & 90°S > \psi > 0° & \text{winter} \\ 0.48 & 30°N > \psi \geq 0° & \text{fall} \\ 0.60 - 0.17\sin 4.5\psi & 90°N > \psi \geq 30°N & \end{cases} \quad (10)$$

We approximate diurnal averaging by multiplying the instantaneous flux by the factor $D(\psi,\delta)$, which is the daylight fraction of the day. δ here specifies the solar declination. An expression for D is obtained from the relation between solar zenith angle, latitude, declination, and local hour angle.

$$\cos\theta = \sin\psi \sin\delta + \cos\psi \cos\delta \cos\Omega t, \quad (11)$$

where $\Omega = 2\pi$ day^{-1}, and t is the time in days measured from local noon. Sunrise and sunset occur at times such that $\cos\theta = 0$, and these times are symmetric about local noon. The daylight fraction is thus

$$\begin{cases} D(\psi,\delta) = \dfrac{1}{\pi}\cos^{-1}[-\tan\psi \tan\delta] \\ D(\psi,\delta) = 0 \text{ if } |\tan\psi \tan\delta| > 1 \quad \text{(polar night)} \end{cases} \quad (12)$$

We note that the mean solar zenith angle to be used in the air mass calculation [Eq. (10)] is obtained from Eq. (12) as

$$\overline{\cos\theta} = \sin\psi \sin\delta + \frac{2}{\pi}\cos\psi \cos\delta. \quad (13)$$

We are neglecting small corrections required in Eq. (12) and Eq. (13) for altitudes $Z > 0$ in the troposphere.

C. Transmissions

The only absorbing gas of consequence for UV radiation, which affects the troposphere, is ozone. The optical depth for ozone absorption is given by

$$\tau_a(\lambda,\psi,Z) = k(\lambda)W(Z)P(\psi,\delta), \quad (14)$$

where $k(\lambda)$ is the ozone extinction coefficient as a function of wavelength λ, $W(Z)$ is the mean global ozone burden above altitude Z, and $P(\psi,\delta)$ is the ratio of ozone burden at latitude ψ during a season specified by the solar declination δ to the mean global ozone burden. For $k(\lambda)$, we adopt the same expression used by SG[2]:

$$k(\lambda) = k_0 \exp[-(\lambda - \lambda_0)/d], \quad (15)$$

with $k_0 = 10$ cm^{-1}, $\lambda_0 = 300$ nm, and $d = 8$ nm. This provides a close fit to measured ozone absorption coefficients over the 280–340-nm region. The ozone

Table I. Downward Glob. l Flux (W/M² nm)

Wavelength (nm)	Solar zenith angle 0°	30°	50°	70°
305	0.70	0.045	0.016	7.9E-4
310	0.172	0.125	0.060	8.5E-3
315	0.296	0.228	0 129	0.032
320	0.414	0.331	0.205	0.069
325	0.515	0.420	0.273	0.108
330	0.602	0.495	0.330	0.142

Note: Ground-level clear-sky solar UV fluxes computed with the algorithms described in the text. The reflected flux has been set to zero and the daylight fraction taken as one to facilitate comparison of these results with those of Shettle and Green.[2]

variation $P(\psi,\delta)$ will not be used in the calculation in Sec. III, Table I, since we wish to compare our calculations with SG who assign various constant ozone burdens. We will simply take $W(Z) \times P(\psi,\delta) = 0.32$ cm atm for our standard calculation presented below.

The optical depth at altitude Z due to molecular scattering is

$$\tau_a(\lambda,Z) = k_a(\lambda) \cdot W_a(Z). \tag{16}$$

We adopt the same expression as SG[2] for the extinction coefficient $k_a(\lambda)$:

$$k_a(\lambda) = k_0 \left(\frac{\lambda_0}{\lambda}\right)^{\nu}, \tag{17}$$

with $k_0 = 0.145\ \text{km}^{-1}$, $\lambda_0 = 300$ nm, and $\nu = 4.27$. Assuming an 8-km scale height for the troposphere, the air burden above Z will be taken as

$$W_a(Z) = 8.423 \exp(-Z/8)\ \text{km atm}, \tag{18}$$

where the 8.423 (total air mass) follows SG.[2] The optical depth for aerosol scattering is again expressed by a function of the form

$$\tau_p(\lambda,\psi,Z) = k_p(\lambda,\psi) W_p(Z), \tag{19}$$

where latitude dependence is introduced through the dependence of the extinction coefficient on predominantly continental or maritime air. From the data of Selby *et al.*[7] we find

$$k_p(\lambda,\psi) = \begin{cases} 0.2 & \psi \geq 30\text{N or } \psi > 70\text{N} \\ -0.56\lambda + 0.47 & 50\text{N} < \psi \leq 70\text{N} \\ -0.42\lambda + 0.37 & 30\text{N} \leq \psi \leq 50\text{N} \end{cases} \tag{20}$$

For the aerosol altitude distribution we again adopt the SG[2] formulation

$$P(Z) = \frac{1.96 \exp(z)}{[0.4 + \exp(z)]^2} + \frac{5.33 \exp(z)}{[2840 + \exp(z/2.5)]^2}, \tag{21}$$

and we compute $W_p(Z)$ as

$$W_p(Z) = \int_Z^{\infty} P(Z)dZ. \tag{22}$$

The transmissions T_{oz}, T_a, and T_p are related to the various optical depths by

$$T_i(Z,\theta) = \exp[-\tau_i(Z) \cdot M(\theta)], \tag{23}$$

where i stands for any of the subscripts oz, a, and p, and $M(\theta)$ is the relative air mass at zenith angle θ. We neglect the slightly different air masses associated with the different attenuating species by SG[2] and take

$$M(\theta) = \left[1 - \frac{\sin^2\theta}{(1 + h/R)^2}\right]^{-1/2} \tag{24}$$

where R is the earth's radius = 6371 km, and $h = 8$ km.

III. Flux Results

Table I shows the result of a sample flux calculation using the algorithms described above. The calculations shown here were programmed on a Radio Shack TRS-80 microcomputer, and many of the initial calculations and tests were programmed on a Texas Instruments SR-52. In order to compare these calculations with SG,[2] we have computed clear-sky fluxes at ground level and arbitrarily set the reflected flux to zero. The daylight fraction was also set to one. For wavelengths ≥310 nm and zenith angles <70°, these fluxes differ from those of SG by ≤10%. At wavelengths ≤305 nm and for a zenith angle of 70° at all wavelengths, larger discrepancies occur with a worst-case difference of 37%. More important than the absolute accuracy of the results is the fact that these calculations exhibit qualitatively correct behavior for the relationships between the various flux components. For example, the ratio of downward diffuse to direct solar flux increases with solar zenith angle. The ratio is 1.4 at the equator for 320-nm radiation and increases to 6 at 60° zenith.

These results encourage us to use the above computationally efficient algorithms for photolysis rate calculations in tropospheric photochemical models. In view of the relatively large uncertainties in the photochemistry, the flux calculations described above will not be the limiting factor in the accuracy of such models. It is the relative computational ease and ample opportunity for sensitivity studies that distinguishes this approach from others.

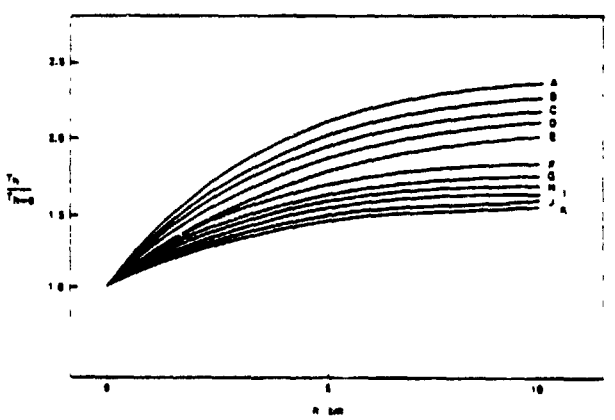


Fig. 1. Variation of total transmission with height h season, and latitude: A—Dec N40°; B—Apr/Sept N60°; C—June S40° D—June N60°; E—Apr/Sept S60°; F—Dec N20° and Apr/Sept N40°; G—June N20°, June S40°, and Dec S60°; H—June 0° and Apr/Sept N20°; I—Apr/Sept S40° and Dec 0°; J—June, Apr/ Sept, Dec S20°, and June N20°; and K—Apr/Sept 0°

Total transmission variation with height, season, and latitude is presented in Fig. 1. Sea-level total transmission is presented in Fig. 2 as a function of season, latitude, and wavelength. In these calculations, the ozone variation $P(\psi,\delta)$ was obtained from the graphical data of Dütsch.[8] Ground albedo was calculated as discussed in Sec. IV.

IV. Photolysis Rate Calculations

For use in zonally and diurnally averaged tropospheric chemistry models photolytic rates were required as a function of latitude, season, and height. These rates were desired in a form that could be readily incorporated in currently existing computer codes for tropospheric chemistry.

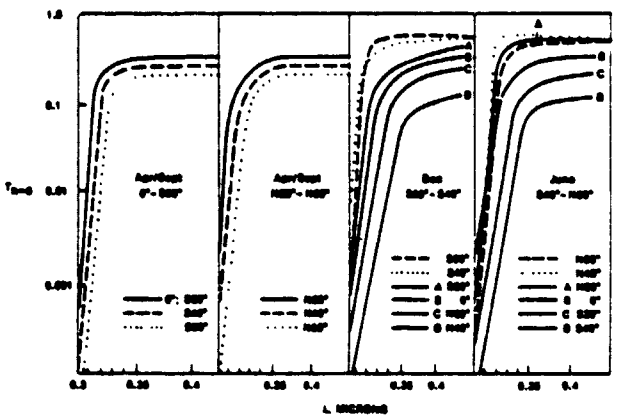


Fig. 2. Total transmission at sea level as a function of season, latitude, and wavelength. The curve shown for Apr/Sept N20° is for April; at 0.3 μm, $T_{h=0}$ is 40% higher in September.

Photolysis rates are the wavelength integrated product of extraterrestrial solar flux, total transmission function, species atomic cross section, and quantum efficiency.[9] For reactions 1b, 2, 4, 5, and 6 of Ref. 10, such information is directly available from Chap. 5 of Refs. 5 and 11. Where possible, the newer values of Ref. 11 were used, with the most recent of Hudson's two absorption cross sections used for reaction R5. Other photolysis rates were calculated as follows. For R1a, we plotted the O_3 visible absorption/O_3 UV absorption vs secz from Table 13 of Ref. 8 and were, therefore, able to use the UV data in Chap. 5 of Ref. 5. For R3a and b, we used the estimate in Ref. 1 for an overhead sun and clear sky, corrected for secz by analogy to Ref. 12 and then corrected only for factors D and f. For R7 and R9, we corrected Table 30 of Ref. 9 for D and f, and, for R8a and b, we corrected the result of Ref. 12 for D and f. In all calculations, ground albedo equals 0.1–0.2, from pp. 444–449 of Ref. 13, and cloud albedo was set equal to 0.52.[13] For wavelength $\lambda > 340$ nm, the flux values of Leighton,[9] rather than Eq. (9), were used.

The resultant photolysis rates are presented in Table II for sea level ($h = 0$). These results can be used for $h \neq 0$ using Fig. 1.

It was gratifying to note that direct integration resulted in J5 = 0.35J2 in Table I, as was assumed in Ref. 10. Also, these photolysis rates vary with height in the same general manner as those of Ref. 4 for low surface albedo. The photolysis rates in Table II are also in rather good agreement with the more approximate rates for some reactions that are tabulated in Ref. 9.

Table II. Diurnally and Zonally Averaged Sea-Level Photolysis Rates (sec^{-1})[a]

Month	Lat. (deg)	J1 (a)	J1 (b)	J(2)	J3 (a)	J3 (b)	J4	J5	J6	J7	J8 (a)	J8 (b)
Apr;Sept	0	2.7/E-4	1.7/E-5	2.7/E-3	1.0/E-2	2.6/E-2	2.2/E-6	9.0/E-4	2.3/E-7	8.3/E-6	1.1/E-3	2.7/E-5
Apr;Sept	S20	2.7	1.4	2.5	0.96	2.5	2.1	8.4	2.1	7.9	1.0	2.7
Apr;Sept	S40	2.4	0.74	2.3	0.78	2.0	1.6	7.6	1.3	6.3	0.8	2.1
Apr;Sept	S60	2.3	0.24	1.9	0.56	1.4	1.1	6.2	0.6	4.1	0.5	1.3
Apr	N20	2.5	1.2	2.4	1.0	2.7	1.9	7.8	1.9	8.5	1.1	2.9
Sept	N20	2.7	1.5	2.4	0.99	2.5	2.0	7.8	2.1	8.1	1.1	2.8
Apr;Sept	N60	1.5	0.08	1.4	0.31	1.3	0.7	4.5	0.4	3.8	0.5	1.1
Apr;Sept	N40	1.7	0.36	2.1	0.71	1.9	1.2	6.5	0.9	5.8	0.8	2.0
Dec	S60	3.2	1.2	3.9	1.5	3.7	2.7	12.6	2.2	11.8	1.5	4.0
Dec	S40	3.5	2.5	3.6	1.6	3.9	2.9	11.4	3.3	12.8	1.7	4.4
Dec	S20	3.3	1.7	3.0	1.3	3.2	2.4	9.5	2.5	10.5	1.4	3.6
Dec	0	2.5	1 3	2.4	0.97	2.5	1.9	8.0	1.9	7.9	1.0	2.7
Dec	N20	2.1	0.61	1.8	0.59	1.4	1.2	5.7	1.0	4.4	0.5	1.5
Dec	N40	1.1	0.01	0.8	0.24	0.6	0.4	2.6	0.2	1.9	0.2	0.5
June	N60	2.0	0.03	3.0	1.2	3.1	1.8	9.5	1.2	10 1	1.3	3.5
June	N40	2.6	1.3	3.5	1.5	3.6	2.5	11.3	2.2	11.9	1.6	4.1
June	N20	2.2	0.95	3.1	1.4	3.5	2.2	10.0	1.8	11.3	1.5	3.9
June	0	1.9	0.68	2.3	0.87	2.2	1.6	7.4	1.3	7.1	0.9	2.4
June	S20	1.6	0.19	1.8	0.59	1.4	1.1	5.7	0.6	4.4	0.5	1.5
June	S40	1.4	0.04	1.0	0.26	0.7	0.5	2.9	0.2	2.1	0.2	0.6

[a] Photolysis rates J1(a), J1(b), etc. correspond to reactions R1(a), R1(b), etc. from Ref. 10. A table entry 3.5 E-5 is identical to 3.5×10^{-5}. Assumed J9 = J7 as in Ref. 10. The relevant reactions are R1(a): $O_3 \rightarrow O_2 + O$; R1(b): $O_3 \rightarrow O_2 + O(^1D)$; R2: $NO_2 \rightarrow NO + O$; R3(a): $NO_3 \rightarrow NO + O_2$; R3(b): $NO_3 \rightarrow NO_2 + O$; R4: $N_2O_5 \rightarrow NO_3 + NO_2$; R5: $HNO_2 \rightarrow OH + NO$; R6: $HNO_3 \rightarrow OH + NO_2$; R7: $H_2O_2 \rightarrow 2OH$; R8(a): $H_2CO \rightarrow HCO + H$; R8(b): $H_2CO \rightarrow CO + H_2$; R9: $CH_3OOH \rightarrow CH_3O + OH$.

This paper was presented during the OSA Topical Meeting on Atmospheric Spectroscopy, Dillon, Colorado, 30 August–1 September 1978.

References

1. R. W. Stewart, S. Hameed, and J. Pinto, IEEE Trans. Geosci. Electron. GE-16, 30 (1978).
2. E. P. Shettle and A. E. S. Green, Appl. Opt. 13, 1567 (1974).
3. M. L. Nack and A. E. S. Green, Appl. Opt. 13, 2405 (1974).
4. F. M. Luther and R. J. Gelinas, J. Geophys. Res. 81, 1125 (1976).
5. A. J. Grobecker, Ed., *CIAP Monograph 1*, U.S. Department of Transportation Report DOT-TST-75-51 (1975).
6. A. J. Grobecker, Ed., *CIAP Monograph 4*, U.S. Department of Transportation Report DOT-TST-75-54 (1975).
7. J. E. A. Selby, E. P. Shettle, and R. A. McClatchey, U.S. Air Force Geophysics Laboratory Report AI GRL-TR-76-0258, Environmental Research Paper 587 (1976).
8. H. V. Dütsch, in *World Survey of Climate and Climatology*, H. E. Landsberg and D. F. Rex, Eds. (Elsevier, New York, 1969), pp. 383–430.
9. P. A. Leighton, *Photochemistry of Air Pollution* (Academic, New York, 1961).
10. R. W. Stewart, S. Hameed, and J. Pinto, J. Geophys. Res. 82, 3134 (1977).
11. R. D. Hudson, Ed., *Chlorofluoromethanes and the Stratosphere*, NASA Reference Publication 1010 (1977).
12. J. G. Calvert, J. A. Kerr, K. L. Demerjian, and R. D. McQuigg, Science 175, 751 (1972).
13. K. Ya. Kondratyev, *Radiation in the Atmosphere* (Academic, New York, 1969).

Paper 32

From the proceedings of the IEEE Conference on Space Instrumentation for Atmospheric Observation, El Paso, TX, April 3-5, 1979.

APPLICATION OF A MICROPROCESSOR CONTROLLED LIDAR TO TROPOSPHERIC OZONE MEASUREMENTS

Richard W. Stewart
Goddard Laboratory for Atmospheric Sciences
NASA Goddard Space Flight Center
Greenbelt, MD 20771

Jack L. Bufton
NASA Goddard Space Flight Center
Greenbelt, MD 20771

Gregory L. Matloff
Systems and Applied Sciences Corporation
Riverdale, MD 20840

SUMMARY

Differential absorption lidar using a dual wavelength, pulsed CO_2 laser and direct detection receiver on board the Space Shuttle is capable of significantly improving the existing global data base on tropospheric ozone burden. A ground based prototype of this lidar is currently under test in the Baltimore-Washington area. The DIAL instrument measures backscattered laser energy at two selected wavelengths in the 9.4 μm band of carbon dioxide. Well known differential absorption algorithms and O_3 absorption line parameters are used along with measured laser energies to compute O_3 concentrations. Data acquisition, recording, and analysis are performed in near real time with a CAMAC microprocessor.

Sensitivity to tropospheric ozone can be obtained in the spaceborne version of this lidar by selecting laser lines located in the wings of the target ozone lines. Substantial absorption occurs only in the troposphere where pressure broadened ozone lines overlap.

Simulation studies[1] using various laser line pairs in the P-branch of the CO_2 9.4 μm band show that the ozone burden retrieval may be weighted to particular altitude regions. The P(22), P(24) line pair is particularly suited to lower tropospheric ozone measurements while other line pairs can be chosen for sensitivity to upper tropospheric or stratospheric ozone amounts. A good first estimate of tropospheric ozone burden can thus be provided by the CO_2 DIAL system using the P(22), P(24) line pair while more accurate values can be deduced from subsequent use of other laser lines or from measurements such as satellite BUV which provide accurate burden measurements of stratospheric ozone.

These simulation studies are based on numerical integration of differences in absorption coefficient profiles at the two selected laser wavelengths using the AFGL absorption line parameter compilations, U. S. Standard Atmosphere ozone profile, and laser software.[2]

Simulation as well as data collection and reduction are performed by an LSI-11 based microprocessor subsystem of the CO_2 lidar. Software control of all these functions provide a flexibility in changing data collection formats and data reduction algorithms.

INTRODUCTION

A carbon dioxide differential absorption lidar using a dual wavelength pulsed laser transmitter and direct detector receiver is capable of space-based burden measurements of several atmospheric species. Potential target constituents are those which have bands in the 9-11 um spectral region and can thus absorb radiation emitted in the 9.6 μm and 10.4 μm bands of the CO_2 laser. The possibility of gaining information on tropospheric ozone is of particular interest since it is involved in most aspects of global-scale tropospheric photochemistry, its source and sink mechanism are a subject of current scientific debate, and an adequate global data base is presently lacking.[3-5]

In this paper we will describe a microprocessor controlled Lidar system currently under construction at the NASA Goddard Space Flight Center and consider some of the problems involved in making space-based measurements of tropospheric ozone.

Paper 33

IEEE TRANSACTIONS ON GEOSCIENCE ELECTRONICS, VOL. GE-16, NO. 1, JANUARY 1978

The Natural and Perturbed Troposphere

RICHARD W. STEWART, SULTAN HAMEED, AND JOSEPH PINTO

Abstract—It is now generally accepted that the troposphere is a region of great chemical complexity and that many human activities may alter the chemical structure of the region. A prerequisite to any realistic assessment of human impacts on pollution and climate is an understanding of the natural budgets of atmospheric gases. This requires a detailed knowledge of physical, chemical, and biological processes within the various reservoirs which are involved in the cycles of these gases. This paper first reviews the processes important in establishing the concentrations of a number of tropospheric species and discusses gaps in our current understanding of these processes. We identify the points at which man may intervene in the major cycles of atmospheric gases and describe the possible consequences of such interventions. Pollutants released into the troposphere may adversely affect the environment by virtue of their chemical interactions with other atmospheric species, their radiative properties, or both. Problems discussed in this review include the growth of atmospheric CO_2 resulting from the burning of fossil fuels and its possible climatic effects, the consequences of increased levels of CO emission on the self-cleansing ability of the troposphere and on the radiation budget, and possible changes in the stratospheric odd nitrogen and ozone amounts due to increased use of fertilizers in agriculture. The magnitude of the perturbations predicted by various model studies are reviewed with particular attention to uncertainties which may affect the results.

I. Introduction

ALTHOUGH problems of urban pollution have been studied for many years, the fact that the global-scale troposphere is an extremely complex chemical system subject to a variety of anthropogenic perturbations is only now being appreciated. Various aspects of the photochemistry of the natural troposphere have been described by several authors [1]-[3]. An understanding of the chemical, biological, and physical processes affecting the concentrations of trace species in clean tropospheric air is clearly a prerequisite to any realistic assessment of the effects of human activities on the state of the atmosphere. Although major problems remain in identifying sources, sinks and chemical reaction paths for some species, it does appear that tropospheric models can be constructed which are consistent with much of our present knowledge of the concentrations of trace species and their chemical reaction rates. At least a qualitative understanding of the impact of industrial activities on the global scale troposphere may be attempted.

Changes in the natural abundance of trace species are of concern for one or both of two reasons. A particular gas may absorb infrared radiation emitted by the earth's surface and thus contribute to the greenhouse effect and the climatic state. Changes in the concentration of such a gas would thus have a direct influence on climate. The spectral region from about 8 to 12 μm is of particular importance since this includes the central portion of the Planck spectrum radiated by the earth and is also a relatively clear "window" in which few natural constituents absorb strongly. Other species have undesirable chemical properties, i.e., they are toxic substances, and still other gases may possess both thermal and toxic properties. It is also possible that a species which is innocuous in itself may, through chemical interactions, alter the concentrations of harmful species. The problems involved in studying the impact of anthropogenic emissions can be quite complex and have many ramifications.

The study of the chemical and climatic effects of industrial emissions into the atmosphere is just beginning and it is clear that a quantitative assessment of these effects requires an understanding of the complex interactions of species within the atmosphere and of the atmosphere with other physical systems such as the oceans, lithosphere, and biosphere. In this paper we will review our current knowledge of the budgets of various species in the natural troposphere and discuss the ways in which man is intervening in these budgets and the possible consequences of such interventions.

II. Abundances of Tropospheric Gases

The concentration of a particular species is determined by competition between various production and loss processes. These processes may consist of physical, biological, or chemical interactions between the atmosphere and other geological systems or they may consist of chemical and physical interactions within the atmosphere itself. With the exception of the noble gases, all elements in the atmosphere undergo exchange processes between the atmosphere and other reservoirs. The sum of these exchange processes constitutes a "cycle" for the element (or species) in question and the average time an atom or molecule spends in a particular reservoir is its lifetime or residence time in that reservoir.

A convenient way of considering the multitude of trace species of importance in the troposphere is to organize them into certain elemental groups as shown in Table I. The first two columns of this table give the name and chemical symbol of various tropospheric species. The third column gives the mixing ratio which for most species must be calculated since their abundances are too small to measure. Those species concentrations which have been measured are indicated by (m) in this column; calculated values were derived by the present authors and represent daytime averaged midlatitude values. The lifetime of a species (column 4) is related to its mixing

Manuscript received July 22, 1977; revised September 30, 1977.

R. W. Stewart is with the NASA Institute for Space Studies, Goddard Space Flight Center, 2880 Broadway, New York, NY 10025.

S. Hameed and J. Pinto are with the Laboratory for Planetary Atmospheres Research, Department of Mechanical Engineering, State University of New York, Stony Brook, NY 11794.

TABLE I
Tropospheric Composition

Species	Chemical Symbol	Fractional Abundance		Lifetime	Thermal IR Band (μm)	Sources	Sinks	Reference
Oxygen Group								
Oxygen (molecular)	O_2	0.2	(m)	10^6 yrs		B	B, P	(87)
	$O_2(^1g)$	10^{-13}		0.26 s		C	C	
Oxygen (atomic)	O	10^{-16}		10^{-3} s				
	$O(^1D)$	10^{-23}		10^{-9} s				
Ozone	O_3	3×10^{-8}	(m)	3 mos	9.6	P, C	P, C	(15)
Carbon Group								
Methane	CH_4	1.4×10^{-4}	(m)	4 yrs	7.7	B	C	(32)
Non Methane Hydrocarbons		$\sim10^{-9}$						(88)
Methyl Radical	CH_3	10^{-21}		10^{-10} s		C	C	
Methoxy Radical	CH_3O	10^{-15}		0.1 s		C	C	
Methylperoxy Radical	CH_3O_2	10^{-11}		10^3 s		C	C	
Methylhydroperoxy Radical	CH_3OOH	10^{-11}		2 d		C	P, C	
Formyl Radical	HCO	10^{-12}		10^{-8} s		C	C	
Formaldehyde	H_2CO	4×10^{-10}	(m)	0.4 d		C, A	C	(87)
Acetaldehyde	CH_3CHO					C	C	
Carbon Monoxide	CO	1.7×10^{-7}	(m)	4 mos		C, A	C, P	(38)
Carbon Dioxide	CO_2	3.3×10^{-4}	(m)	3 yrs		B, A	B, P	(43)
Nitrogen Group								
Nitrogen	N_2	0.8	(m)	10^6 yrs		B, P	B, P	(87)
Nitrous Oxide	N_2O	3.3×10^{-7}	(m)	20 yrs?	7.8, 17.	B, P	C	(58)
Nitric Oxide	NO	10^{-11}				A, B	C	
Nitrogen Dioxide	NO_2	10^{-10}	(m)		7.6	A, B, C	C	(20)
Nitrogen Trioxide	NO_3	10^{-14}		15 s		C	C	
Dinitrogen Pentoxide	N_2O_5	10^{-14}		15 s		C	C	
Nitrous Acid	HNO_2	10^{-12}		10^3 s		C	C	
Nitric Acid	HNO_3	10^{-9}		1 week	5.9, 7.5 11.3, 21.8	C	P, C	
Ammonia	NH_3	10^{-9}	(m)	1 week	10.33	B, A	C, P	(63)
Amine Radicals	NH_2	10^{-21}		10^{-5} s		C	C	
Aminoxy Radical	NH_2O	10^{-16}		1 s		C	C	
Aminoperoxy Radical	NH_2O_2	10^{-12}		1 min		C	C	
Nitrosyl Radical	NHO	10^{-14}		10 s		C	C	
Ammonium Nitrate	NH_4NO_3	10^{-10}		1 week		C	P	
Hydrogen Group								
Hydrogen (molecular)	H_2	5×10^{-7}	(m)	10 yrs		C, B, A	C	(77)
Hydrogen (atomic)	H	10^{-21}		10^{-7} s		C	C	
Water	H_2O	0.014	(m)	1 week	6.25, 12.0, 20.0, 10^{-}	P	P	(87)
Hydroxyl Radical	OH	10^{-14}	(m)	1 sec		C	C	(34) (33)
Hydroperoxyl Radical	HO_2	10^{-11}		4 min		C	C	
Hydrogen Peroxide	H_2O_2	10^{-9}		2 d		C	C	
Sulfur Group								
Hydrogen Sulfide	H_2S	2×10^{-10}	(m)	3 d		B, A	C	(78)
Dimethyl Sulfide	$(CH_3)_2S$					B	C	
Dimethyl Disulfide	$(CH_3)_2S_2$					B	C	
Sulfur Dioxide	SO_2	2×10^{-10}	(m)	1 week	8.7, 7.3	B, A	C, P	(78)
Carbonyl Sulfide	COS	2×10^{-10}				B	C	
Thiyl Radicals	HS	10^{-21}		10^{-4} s		C	C	
Sulfoxy	SO	10^{-18}		10^{-3} s		C	C	

TABLE 1 (*continued*)

Species	Chemical Symbol	Fractional Abundance	Lifetime	Thermal IR Band (μm)	Sources	Sinks	References
Sulfur Trioxide	SO_3	10^{-12}	10^{-6} s		C	C	
Sulfurous Acid	H_2SO_3	10^{-10}	1 week		C	P	
Sulfuric Acid	H_2SO_4	10^{-10}	1 week		C	P	
Halogen Group							
Trichlorofluoromethane (Freon 11)	$CFCl_3$	10^{-10} (m)	80 yrs	9.22, 11.82	A	C	(89)
Dichlorodifluoromethane (Freon 12)	CF_2Cl_2	3×10^{-10} (m)	100 yrs		A	C	(89)
Dichlorofluoromethane (Freon 21)	$CHFCl_2$	1.4×10^{-11} (m)	2 yrs		A	C	(36)
Chlorodifluoromethane (Freon 22)	CHF_2Cl		16 yrs		A	C	
Carbon Tetrachloride	CCl_4	10^{-10} (m)	60 yrs	12.99	A, B	C	(89)
Methyl Bromide	CH_3Br	4.7×10^{-12} (m)	1.5 yrs	16.4, 10.5, 7.66, 6.92	A, B	C	(36)
Methyl Chloride	CH_3Cl	7×10^{-10} (m)	1.4 yrs	13.66 9.65, 7.14	A, B	C	(89)
Methyl Iodide	CH_3I	10^{-11} (m)		10.76, 11.36 7.99, 6.94	A, B	C	
Hydrogen Chloride	HCl	10^{-9}	1 week		A, P	C	(89)
Chloroform	$CHCl_3$	9.4×10^{-12} (m)	7 mos		A, B	C	(36)
Trichloroethylene	C_2HCl_3	$\sim2\times10^{-10}$ (m)	6 d		A	C	(36)
Tetrachloroethylene	C_2Cl_4	3.1×10^{-11} (m)	3 mos		A	C	(36)
1,1,1 Trichloroethane (Methyl Chloroform)	CH_3Cl_3	7×10^{-11} (m)	3.3 yrs		A, B	C	(36)
Vinyl Chloride	C_2H_3Cl		2 mos		A	C	
Phosgene	$COCl_2$	2×10^{-11} (m)			A	C	(36)
Noble Gases							
Argon	A	9.3×10^{-3} (m)			P		(90)
Neon	Ne	1.8×10^{-5} (m)			P		(90)
Krypton	Kr	1.1×10^{-6} (m)			P		(90)
Xenon	Xe	9.0×10^{-8} (m)			P		(90)
Helium	He	5.2×10^{-6} (m)			P	P	(90)
Radon	Rn		3 d		P	P	

ratio in that those which are highly reactive and have short lifetimes will generally have small abundances. Absorption bands in the thermal IR are indicated in column 5. Sources and sinks of the various species are discussed below; columns 6 and 7 of Table I note whether the sources and sinks for a particular species are primarily anthropogenic (A), biological or microbiological (B), chemical reactions in the atmosphere (C), or physical (P).

III. The Oxygen Group

The most abundant member of this group, molecular oxygen (O_2), is a major constituent comprising about 20 percent of the atmosphere. The oxygen cycle has been reviewed by Walker [4] and by several others [5]-[7] in the context of discussions of the carbon cycle. To address the question of intervention by human activities in the oxygen cycle we must note that there are two different time scales in which oxygen is exchanged between the atmosphere and terrestrial reservoirs. The shorter of these time scales involves the exchange of about 3×10^5 Mt/yr of oxygen between the atmosphere and biosphere via processes of photosynthesis and respiration and decay. The reservoir of atmospheric oxygen involved in this exchange (10^9 Mt oxygen) is much larger than the reservoir of biospheric carbon (3.5×10^6 Mt) or of available fossil fuels (7×10^6 Mt). Since 2.67 Mt of O_2 will oxidize 1 Mt of carbon it is clear that human intervention in the carbon cycle, either through nearly complete combustion of available fossil fuels or through substantial alterations in the populations of photosynthesizing organisms, will have small (less than 4 percent) impact on the atmospheric O_2 abundance over times of the order of hundreds of years.

The longer time scale involves the exchange of O_2 between the atmosphere and crustal sediments. The source of oxygen on this scale is due to about 300 Mt/yr of O_2 produced by photosynthesis which are not utilized in the oxidation of dead organic carbon. The stoichiometric equivalent amount of carbon is sequestered in ocean sediments. A corresponding sink of 300 Mt/yr for O_2 results from erosion and weathering of these sediments as they are again exposed to the atmosphere. The reservoir of sedimentary carbon contains about 5.5×10^{10} Mt and could readily absorb all the O_2 presently in the atmosphere if the O_2 source were extinguished. The time over which this would occur is about 10^9 Mt/(300 Mt/yr) or 3 million years. Because of the long time scale involved in

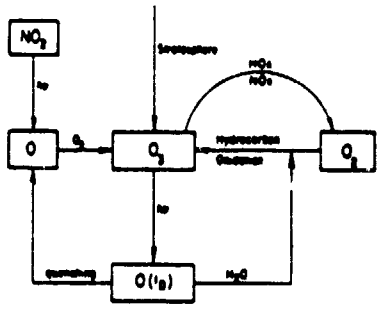


Fig. 1. Oxygen budget of the troposphere. The three sources of ozone are downward transport from the stratosphere, NO_2 photolysis, and hydrocarbon oxidation. The dominant chemical sinks due to reaction of ozone with odd hydrogen and nitrogen radicals and of $O(^1D)$ with water vapor are shown. The lesser sink due to reaction at the ground is not indicated.

potential human impacts on the oxygen cycle, we will not discuss it in detail but will turn to some problems of more immediate concern involving other members of the oxygen group.

Ozone is the next most abundant member of the oxygen group and is a chemically important tropospheric constituent. Almost all of tropospheric photochemistry is initiated by the photolysis of ozone in the 300-310-nm region of the ultraviolet spectrum. The reaction is

$$O_3 + h\nu \longrightarrow O_2 + O(^1D).$$

The $O(^1D)$ thus formed may then react with water vapor

$$O(^1D) + H_2O \longrightarrow 2OH$$

giving two of the highly reactive hydroxyl radicals which in their turn initiate several complex reaction chains. Another potential precursor of tropospheric photochemistry is NO_2 which can dissociate over the UV wavelength range from 300-400 nm.

$$NO_2 + h\nu \longrightarrow NO + O$$

with subsequent involvement of both O and NO in chemical reactions.

The oxygen budget of the troposphere is represented schematically in Fig. 1. This figure indicates that NO_2 photolysis and downward transport from the stratosphere are primary sources of tropospheric ozone. Photolysis of this ozone results in the production of $O(^1D)$ and, after reaction of this species with H_2O, in OH radicals. These in turn oxidize methane and other hydrocarbons to produce more ozone. Ozone is lost via reactions with NO_x (NO and NO_2) and HO_x (OH and HO_2) radicals.

A problem that has generated much recent debate is that of the sources and sinks of tropospheric ozone. The "classical" view has been that ozone is transported downward from the stratosphere, where it is created as a result of O_2 dissociation by UV wavelengths less than 240 nm, and is destroyed at the ground [8]. The observed background concentration is about 3 pphm and if downward transport and ground-level destruction are in fact the major source and sink mechanisms for tropospheric ozone, its lifetime is of the order of one month. Chameides and Walker [9], [10] proposed an alternative view that ozone in the lower troposphere is photochemically created, especially as a consequence of the methane oxidation chain. The role of reactive hydrocarbons in ozone production in polluted urban air is well known [3], [11] and the possible importance of the much less reactive but more abundant methane in the global scale production of ozone was noted by Crutzen [12].

The details of the reaction scheme by which methane oxidation results in ozone production are discussed in Fishman and Crutzen [13]. The net effect, which is the sum of 15 individual reactions, is

$$CH_4 + 6O_2 \longrightarrow H_2O + CO_2 + H_2 + 3O_3.$$

This reaction sequence essentially extracts O_2 molecules from the ambient reservoir and makes their component atoms available for ozone formation. It thus provides an alternative to the direct splitting of O_2 by ultraviolet radiation which does not occur in the troposphere.

The importance of transport in establishing the tropospheric ozone abundance is supported by observational evidence on the latitudinal distributions of this constituent. The tropospheric ozone burden (total amount of ozone in a column) shows maxima in regions of enhanced troposphere-stratosphere exchange, 30, 45, and 60° latitude [14], [15]. Photochemical models have not addressed the problem of the observed latitudinal distribution of ozone, but there is no obvious photochemical argument which might account for it.

The sharp disagreement between the conclusions of the photochemical models of Chameides and Walker [9], [10] and the arguments supporting a purely dynamic source for tropospheric ozone have motivated additional model studies [16], [17]. The results of these more recent calculations indicate that photochemical production alone cannot account for observed levels of tropospheric ozone and that, in fact, the troposphere acts as a net chemical sink. In the calculation of Chameides and Stedman [16], 70 percent of the ozone in a typical midlatitude tropospheric column is produced by downward transport from the stratosphere and 30 percent is photochemically produced. In this typical column 70 percent of the ozone is chemically destroyed and 30 percent destroyed by reaction at the ground. Calculations made with the box model of Stewart *et al.* [17] lead to similar conclusions regarding the relative importance of transport and photochemistry in establishing the tropospheric ozone distribution.

The divergent conclusions of the newer and older photochemical models illuminate the principal problem in state-of-the-art models of tropospheric composition, which is their sensitivity to the data base used in the calculations. The most important component of this data base is the large set (usually $\gtrsim 60$) of reaction rates for the chemical reactions used in the calculations. These are constantly being revised as new experimental data become available. The most generally accepted chemical data base is that provided by the compilation of Hampson and Garvin [18] in which these authors, and other cited in their publication, review the available experimental data and provide preferred values for a large body of

chemical reaction rates. This review process is by its nature always somewhat dated, and there are usually more recent data of interest which must be considered in model calculations.

The different conclusions reached by the older [9], [10] and newer [16], [17] models of tropospheric ozone may be traced for the most part to the faster rates adopted for certain important reactions in which odd hydrogen and odd nitrogen radicals combine, e.g.,

$$NO_2 + OH \xrightarrow{M} HNO_3.$$

These rates were revised between publication of the models referred to above. Such reactions reduce the concentration of OH in the troposphere and thus render the methane oxidation mechanism less effective.

Another factor leading to differing conclusions in these models is the different NO_2 background adopted. The earlier models used 3-ppb NO_2, a value suggested by Robinson and Robbins [19] based on their analysis of the nitrogen budget and the sparse data then available. As additional data have become available it has become clear that 3 ppb is too high to be representative of global background NO_2 values. The spectroscopic measurements of Noxon [20] in particular suggest that $NO_2 \lesssim 0.1$ ppb. The use of lower NO_2 values ($\lesssim 0.5$ ppb) in the later photochemical models reduces ozone production resulting from NO_2 photolysis.

The newer photochemical models discussed above represent a synthesis of the arguments concerning photochemical and dynamic sources for tropospheric ozone. They establish the importance of photochemical interactions in contributing to the overall ozone abundance, but are consistent with the view that transport plays a major role in establishing the global ozone distribution. Unfortunately, this consensus has proved shortlived. A recent measurement of the rate of the reaction

$$HO_2 + NO \longrightarrow OH + NO_2$$

by Howard and Evenson [21] indicates that it is 8 to 20 times faster than the previously accepted values. This has extremely important consequences for the problem of tropospheric ozone and for tropospheric chemistry generally, since this reaction produces NO_2, an ozone precursor, and OH, which initiates the methane oxidation chain and again produces ozone.

In the model of Stewart *et al.* [17], to maintain the average observed tropospheric background value of 4.5-pphm O_3, the troposphere must import 1280 Mt/yr of ozone from the stratosphere. This figure is consistent with transport values quoted in the meterorological literature [15], [22]. However, using the Howard and Evenson value for the reaction given above in the Stewart *et al.* model, the troposphere switches from being a strong chemical sink to a strong chemical source; it must export 2000 Mt/yr of ozone. This is an unacceptable result because it is inconsistent with the observed tropospheric distribution as explained previously, and because the observed troposphere-stratosphere ozone gradient implies downward rather than upward transport.

Thus the problem of the sources and sinks of tropospheric ozone remains. Even if revision of other reaction rates is found to provide a chemical sink which compensates for the chemical source given by this rate, ozone may be found to have too short a residence time to be consistent with the indicated importance of transport in establishing its distribution.

IV. The Carbon Group

A. Natural Carbon Cycle

The principal carbon containing species in the atmosphere are CO_2, CH_4, and CO, which have mean concentrations of about 330 ppm, 1.4 ppm, and 0.12 ppm, respectively. Other hydrocarbons such as aldehydes, olefins, alkanes, and terpenes are generally present in much smaller amounts. The global scale effects of the carbon group on pollution and climate have centered on the three species noted above. Higher molecular weight hydrocarbons and CO are important also in urban scale pollution problems [11], [23], [24]. Carbon dioxide and methane are climatologically significant gases since they absorb in the thermal IR. The contribution of CO_2 to the "greenhouse effect" is well known, but methane has a band at 7.7 μm, and changes in CH_4 abundance are also potentially important. Carbon monoxide is not active in the thermal IR but is of concern due to its chemical properties. In high concentrations CO is toxic, but on the global scale our concern is centered more on its possible chemical interaction with and impact on species such as OH and CH_4 [25], [26].

The abundance of carbon dioxide is regulated by interaction between the troposphere and biosphere [27] via the process of photosynthesis which we may write schematically as

$$CO_2 + H_2O \rightleftharpoons CH_2O + O_2.$$

The forward reaction represents the utilization of CO_2 by plants in the synthesis of organic compounds here represented by CH_2O. The reverse reaction represents the return of CO_2 to the atmosphere by oxidation of dead organic matter. The flow of carbon between the atmosphere and other reservoirs is shown in Fig. 2. Several quantitative models of the carbon cycle have been developed [5], [6], [28]-[30] which may differ from one another and from Fig. 2 in the details of the adopted reservoirs and fluxes.

In Fig. 2 we have given the consumption of CO_2 by the biosphere as 1.5×10^5 Mt/yr, 1.1×10^5 by the land biomass and 4×10^4 by the oceanic biomass. These numbers, like most of the flux and reservoir values of Fig. 2, are meant to be representative, but widely varying values can be found in the literature. The exchange of CO_2 between the atmosphere and biomass by photosynthesis is balanced to within one part in a thousand. The slight imbalance between consumption and release of CO_2, represented by the sedimentation flux in Fig. 2, has resulted in the deposit over geologic time of a reservoir of fossil organic carbon of perhaps 1.2×10^7 Mt (31), of which about 7.3×10^6 Mt is available for exploitation.

Methane and carbon monoxide are less abundant but important members of the carbon group. Methane is approximately uniformly distributed in the troposphere with a con-

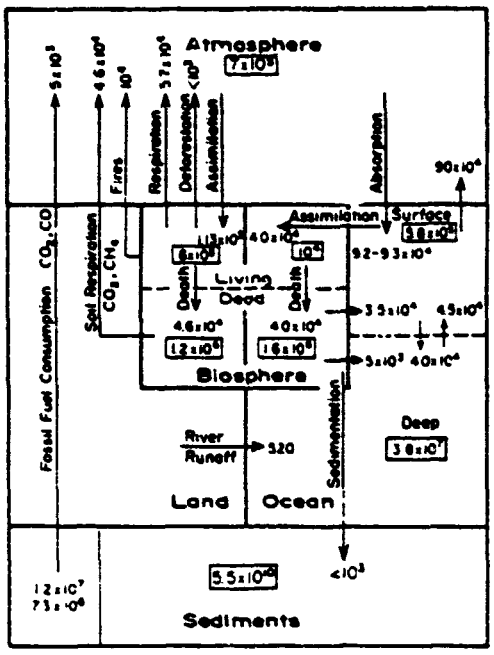


Fig. 2. The Carbon cycle. The number of megatons (Mt) of carbon in the indicated reservoirs are shown in boxes. The total fossil fuel carbon (1.2×10^7 Mt) and the amount available for use (7.3×10^6 M) are shown at the left of the sediment reservoir. The fluxes between reservoirs are shown by the arrows and given in Mt (C)/yr. The physical or biological process giving rise to these fluxes is stated next to the arrow in most cases.

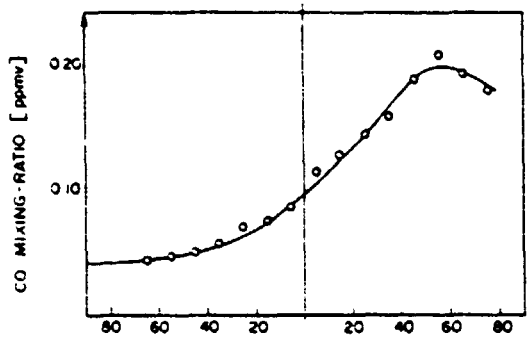


Fig. 3. Latitudinal variation of CO in the troposphere (after Seiler [38]).

centration of 1.4 ppm [32]. At least eighty percent of tropospheric methane is produced by the anaerobic decomposition of organic matter in such places as swamps and paddy fields, and its sources are thus biological and microbiological. The remainder of the CH_4 is output from natural gas wells. A present estimate of the total source strength of CH_4 is about 10^3 Mt/yr [32] and is included in the soil respiration flux in Fig. 2. There are 4×10^3 Mt of CH_4 in the atmosphere which, combined with the above source strength, implies an atmospheric residence time of 4 years. It is now generally believed that oxidation of methane via the reaction

$$CH_4 + OH \longrightarrow CH_3 + H_2O$$

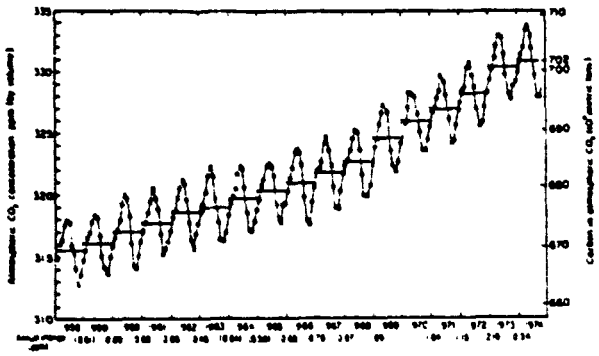


Fig. 4. Monthly average values of CO_2 measured at Mauna Loa Observatory, Hawaii (after Baes *et al.* [31]). The seasonal oscillation of CO_2 is shown superposed upon the long term increase due to industrial activity. Horizontal bars indicate mean annual concentrations.

provides the major sink. The effectiveness of this sink depends on the abundance of tropospheric OH radicals and on the rate of the above reaction. The mean tropospheric abundance of OH is not well established. Measurements [33]-[35] show it to be quite variable and theoretical arguments support OH concentrations ranging from 2×10^5 to 3×10^6 molecules/cm^3 [36], [37]. Adopting a rate of $k = 4.6 \times 10^{-15}$ cm^3 s^{-1} for the CH_4 + OH reaction [18] gives a lifetime $\tau_{CH_4} = (k[OH])^{-1}$ ranging from 2.3 to 35 years. There is thus at least a factor of 10 uncertainty in the parameters determining the methane abundance.

Carbon monoxide is approximately uniformly distributed vertically in the troposphere but shows a strong latitudinal gradient. The globally averaged concentration of CO is about 0.12 ppm, but there is a factor of four difference in mean concentrations between the northern (0.20 ppm) and southern (0.05 ppm) hemispheres [38]. (See Fig. 3.) This interhemispheric concentration difference indicates the importance of northern hemisphere anthropogenic sources, mainly automobile and space heating, in determining the global CO abundance, but the oxidation of methane provides a strong natural production and the relative importance of these two mechanisms is still a matter of debate [39]-[41].

B. Perturbations to the Carbon Cycle

The long-term increase of CO_2 is the best documented of the global scale changes in the atmospheric environment that have occurred as a result of human activities [42], [43]. Fig. 4 (after Baes *et al.* [31]) shows monthly average values of CO_2 concentrations measured at Mauna Loa Observatory, Hawaii. The long-term increase is quite apparent. Seasonal variations in photosynthesis produce the annual cycle. Carbon dioxide absorbs in the thermal IR, and it has been suggested by several theoretical studies that substantial increase (e.g., doubling) of the CO_2 amount will lead to climatologically significant temperature increases in the lower atmosphere. There are two parts to a quantitative evaluation of the environmental impact of enhanced CO_2 levels: the first is to project future CO_2 concentrations as a function of time, the second is to calculate the change in surface temperature T_s

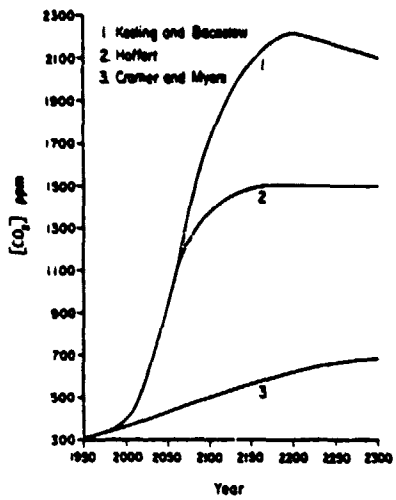


Fig. 5. Projected atmospheric CO_2 increases due to fossil fuel burning from three model studies. The production scenarios are similar in these studies and the differing results chiefly reflect differences in the treatment of the oceanic CO_2 sink.

due to these concentrations. Each part of this problem has been addressed by several investigators but with no consensus on either point.

The long-term increase of CO_2 has been modeled in several studies [28], [29], [44], [45], and the results of some of these are shown in Fig. 5. Projections of future CO_2 levels in these model studies are obtained by solving the continuity equation for CO_2 which equates the rate of change of CO_2 in a specified reservoir to the difference between sources and sinks for that reservoir. The models shown in Fig. 5 use comparable source functions for the anthropogenic emission (fossil fuel burning). Hoffert [29], for example, fits a Gaussian source function to historical emission data over the 1950–1970 period. This source function results in consumption of about 70 percent of available fossil fuels in the hundred-year period centered about the year 2069. The antropogenic source functions used in the other two models of Fig. 5 differ in some details but are essentially similar in magnitude and functional form.

The two central issues in current research on this problem involve the roles of the oceans and the biomass as sources and sinks of CO_2. The substantial difference in the long range projections of the authors cited in Fig. 5 are due to the differing treatments of the oceanic sink in their models. Hoffert [29] has discussed the differences between his model and that of Cramer and Myers [28] as being primarily due to treatment of the deep-ocean CO_2. In Hoffert's model this is determined from the carbonic acid equilibrium

$$CO_2 + H_2O \rightleftharpoons H^+ + HCO_3^-$$

$$HCO_3^- \rightleftharpoons H^+ + CO_3^=$$

which constrains the deep-ocean CO_2 to smaller values than in the Cramer and Myers model in which this CO_2 is determined by exchange with the mixed layer and the ocean floor. The ability of the oceans for CO_2 uptake is smaller in Hoffert's model and thus a greater amount of CO_2 accumulates in the atmosphere. In the Keeling and Bacastow model [45] the thermal stability of the main oceanic thermocline (70–1000 m deep) prevents the complete mixing of CO_2 between this reservoir and the warm surface layer, thus rendering the ocean a still less effective sink than in the Hoffert model.

Another issue which has recently been raised [46], [47] concerns the effect of human alterations of the land biosphere such as deforestation and agricultural expansion on atmospheric CO_2 changes. Bolin [47] suggests that such changes in the land biota constitute a source of CO_2 equal to about 10–35 percent of the emission due to fossil fuel combustion. He also concludes that since this auxiliary source of CO_2 has previously gone unrecognized the oceans must provide a more effective sink that hitherto realized. The results of Adams *et al.* [46] and Bolin [47] imply that the oceans are capable of absorbing 58–83 percent of the fossil fuel CO_2 emitted into the atmosphere. This conclusion is in sharp contrast to the trend of recent model development in which the oceans play a diminished role from that of earlier models. Broecker [48], for example, estimates that of the CO_2 emitted into the atmosphere, 52 percent remains in the atmosphere, 38 percent dissolves in the oceans, and 10 percent is assimilated by the biosphere. Thus there is no current consensus as to the magnitude of the oceanic sink or whether the land biota constitute a source or a sink of CO_2.

Part of the difficulty in assessing the role of the biosphere in the carbon cycle stems from the fact that relatively small changes in biomass would substantially alter the CO_2 flux between atmosphere and biosphere. The deforestation flux of 1000 Mt/yr shown in Fig. 5 corresponds to a decrease in the land biosphere of less than 0.2 percent. A similarly small increase would provide a sink of the same magnitude. As we have noted earlier, the ocean biomass is controlled by nutrients other than carbon and we shall return to this point in our discussion of the nitrogen cycle.

The thermal effects of a given CO_2 increase have been studied by several authors [49]–[52], and a review of these efforts through 1975 has been published by Schneider [53]. Differences in predicted surface temperature increase ΔT_s among various 1-D radiative-convective models are fairly well understood in terms of differing assumptions and computational procedures. The most detailed radiative transfer calculation appears to be that of Augustsson and Ramanathan [54] who included the temperature dependence of the band absorptance in the 15-μm bands as well as the contribution to the greenhouse effect of the weak bands in the 12–18-μm, 9–10-μm, and 7–8-μm regions. Their result for a doubling of CO_2 concentration is a ΔT_s of 1.98 K. Manabe and Wetherald [50] have utilized a general circulation model to calculate a ΔT_s of 2.9 K (globally averaged) some 46 percent larger than the value of Augustsson and Ramanathan. In the Manabe and Wetherald model the surface temperature increase is greater at high latitudes due to greater thermal stability of the troposphere and to decreased albedo resulting from recession of the snow boundary. This relatively large surface temperature increase at high latitudes results in a larger globally averaged ΔT_s than is obtained with radiative-convective models.

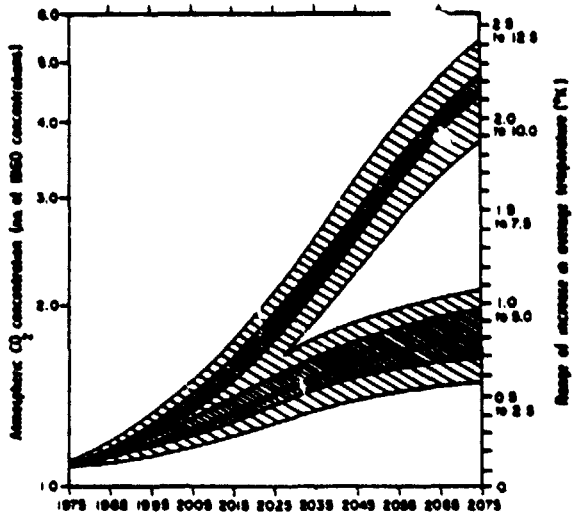


Fig. 6. Projected growth in atmospheric CO_2 and associated range of ΔT_s values for two scenarios of fossil fuel utilization (after Baes *et al.* [31]). These curves show the sensitivity of the model described by Baes *et al.* to high and low use production scenarios (inner solid lines), assimilation of CO_2 by land biota (inner band), and oceanic uptake (outer band). The right hand vertical axis shows the range of uncertainty in ΔT_s for each value of CO_2 increase.

The greatest uncertainty in model calculations of surface temperature response to increased CO_2 abundance is in the treatment of other meterorological parameters such as relative humidity and cloud amount, which may also respond to temperature changes and feed back into the CO_2 perturbation. Augustsson and Ramanathan [54], for example, show differences of up to 62 percent in ΔT_s between models with constant cloud top altitude and others with constant cloud top temperature.

According to Schneider [53], a present estimate of ΔT_s resulting from a doubling of atmospheric CO_2 abundance would be $1.5\ \text{K} < T_s < 3\ \text{K}$, but, as Schneider cautions, this estimate could be considerably modified by the effects of important feedback mechanisms not included in current models.

Considering the combined uncertainties of projecting future CO_2 levels and of computing ΔT_s for a given CO_2 increase, it does not seem reasonable at present to attempt a quantitative prediction of temperature increases resulting from CO_2 emissions at any specified future time. The uncertainties in the CO_2 and ΔT_s projections are indicated in Fig. 6 (after Baes *et al.* [31]). The higher of the two solid lines represents an assumed 4.3 percent annual growth rate in fossil fuel burning reduced in proportion to the fraction of the net supply that has been used; the lower of the two solid lines represents an assumed 2-percent growth rate until 2025 followed by a symmetrical decrease. The inner band in each of these cases represents the effect of a ±2000 Mt/yr variation in the net flux to the land, while the outer band represents the effect of varying the assumed ocean uptake from 40 to 60 percent. The right hand vertical axis shows the range of uncertainty in ΔT_s (as estimated by Baes *et al.*) for each value of CO_2 increase.

Another perturbation to the carbon group which has been discussed recently involves possible alteration of the CH_4 concentration as a result of changes in the anthropogenic CO flux into the atmosphere [25], [25]. According to calculations by Sze [25], a CO emissions increase of 4 percent per year from 1940 to 1971 followed by a 1.33 percent per year increase thereafter will result in a doubling of CH_4 concentration over its 1950 value by the year 2035. Chameides *et al.* [26] consider a 4.5-percent increase in CO emissions from 1976 and find an increase of CH_4 from present levels to about 2.45 ppm by the year 2010 (a 75-percent increase).

These projected CH_4 increases occur because CO is an effective scavenger of the hydroxyl radical through the reaction $CO + OH \rightarrow CO_2 + H$, and OH is the major species capable of initiating the methane oxidation sequence. Thus more CO results in less OH and this, in turn, causes an increase in CH_4 amount.

Hameed *et al.* [55] have pointed out that the magnitude of the CH_4 perturbation resulting from a given CO flux increase is extremely sensitive to the NO_x background assumed in the model and, since this at present is only poorly known, quantitative statements concerning the magnitude of the CO-CH_4 perturbation are not yet feasible. According to the calculations of Hameed *et al.*, a doubling of the present level of CO emissions would result in an increase of CH_4 to 1.9 ppm for $NO_x = 0.1$ ppb. The reason for this sensitivity is that NO_x compounds are also effective scavengers of OH through reactions such as

$$NO_2 + OH \xrightarrow{M} HNO_3.$$

The hydroxyl radical concentration is thus much more sensitive to the CO flux for low values of NO_x than for higher values.

According to *Wang et al.* [56], a doubling of CH_4 concentration would result in an incremental surface temperature ΔT_s of 0.28 or 0.40 K for a fixed cloud-top temperature model depending on whose absorption data is used. It is important to more firmly establish the possible magnitudes of methane increases and to study in greater detail the more general changes which would result from lowered hydroxyl radical concentrations.

V. The Nitrogen Group

A. Natural Nitrogen Cycle

Nitrogen as N_2 is the most abundant gas in the atmosphere, but it is relatively inert chemically and therefore has little direct interaction with other tropospheric species. Other forms of tropospheric nitrogen such as ammonia (NH_3), nitric oxide (NO), and nitrogen dioxide (NO_2) result from various microbiological and industrial processes which break the N_2 bond. Such processes are said to "fix" nitrogen in a biologically useful form. These fixation processes occur naturally in various soil and marine organisms and to a much lesser extent by the reaction of N_2 and O_2 in lightning discharges. Industrial production of fertilizer is the principal source of artificial fixation and a debate is in progress over the degree to which this process may upset the natural cycling of nitrogen through the ecosystem. The current status of this problem will be described below.

After N_2, nitrous oxide is the most abundant form of nitro-

gen in the troposphere. It appears to be relatively uniformly distributed spatially, but temporal fluctuations have been interpreted as implying an atmospheric residence time from 4 to 70 yr for this gas [57], [58]. Chemical destruction in the stratosphere alone would result in a longer lifetime–118 to 160 yrs [59]. With an atmospheric abundance from 0.25 ppm to 0.30 ppm (1800 to 2160 Mt as N_2O), this range of lifetimes implies source and sink strengths ranging from 11.3 Mt/yr to 540 Mt/yr. Although the source of atmospheric N_2O is clearly denitrifying bacteria, there is still a problem in identifying a sink of the appropriate magnitude. The role of the oceans in this regard is a subject of debate. Hahn [60] identified the oceans as a possible net source of N_2O, whereas McElroy *et al.* [61] believe that a flux of N_2O into the oceans may exist and provide the necessary sink. More data are required to accurately assess the role of the oceans in the nitrogen cycle.

Nitrous oxide is important because of its fundamental role in controlling the abundance of stratospheric ozone. In the stratosphere N_2O dissociates by reaction with $O(^1D)$ to form nitric oxide which catalytically reduces the ambient ozone. N_2O also has an absorption band at 7.78 μm, and it has been estimated that a doubling of its concentration could result in a global mean increase of 0.68 K in surface temperature [56]. Due to the strength and position of the N_2O 7.78-μm band it is the second most important absorber of thermal IR (after CO_2 and neglecting the highly variable water vapor). The gas is of potential climatological significance due to both its thermal IR and chemical properties.

Ammonia, the next most abundant form of atmospheric nitrogen, exhibits strong spatial and temporal variabilities [62]. There is a strong decrease of NH_3 with altitude, the concentration above 2 km being about a third of the ground level value. Seasonally, mean ground level NH_3 concentrations may vary from about 6 ppb in summer to 2 ppb in winter over land areas, but over the oceans there is a sharp decrease from about 0.2–1.0 ppb. The natural source of atmospheric NH_3 is bacterial decomposition of biologically fixed nitrogen found in plant tissues. The magnitude of this production depends on soil type, temperature, pH and, moisture content. Dawson [63] has recently calculated a soil source of 47 Mt/yr of NH_3 from unperturbed land on a global basis. This would be balanced by corresponding rainout and dry deposition. The atmospheric residence time of NH_3 is about a week before it is removed by precipitation.

An anthropogenic source of ammonia results from the volatization of ammonium contained in nitrogeneous fertilizers applied to agricultural soils. The remainder of the fertilizer nitrogen is assumilated by the biomass or undergoes transfer from the inorganic nitrogen soil pool to other reservoirs by processes noted below. The quantitative relationship between these processes is not established, although it probably depends on the same soil variables described previously [63]. Ammonia is thus strongly coupled to the processes affecting the cycling of nitrogen through the ecosystem via its interactions in the biosphere.

The oxides of nitrogen, NO and NO_2, are the least abundant tropospheric nitrogen species. Their global distribution and background concentration are poorly known in the troposphere. Recent measurements by Noxon [20] indicate values of the order of 0.1 ppb or less. Tropospheric sources for NO_x are combustion processes and lightning (10–40 Mt/yr) [64], [65], and denitrification of soil nitrogen. NO_x plays an important role in tropospheric chemistry involving the production and loss of ozone [17] and the effect of CO emissions on CH_4 abundances [55]. Its global mean background concentration must be much better established.

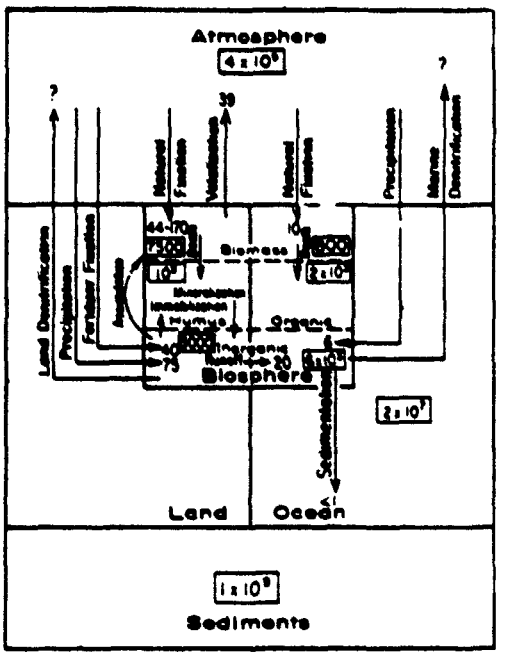


Fig. 7. The Nitrogen cycle. The number of Mt nitrogen in the indicated reservoirs are shown in boxes. The fluxes between reservoirs are shown by the arrows and given in Mt (N)/yr. The physical or biological process giving rise to these fluxes is stated next to the arrow.

The nitrogen cycle is schematically represented in Fig. 7. The reservoir values are similar to those stated in published reviews [7], [66] but have been adjusted to be consistent with observed C/N ratios of 80/1 for the living terrestrial biomass and 10/1 for other biospheric reservoirs.

B. Perturbations to the Nitrogen Cycle

Man's intervention in the nitrogen cycle is largely through the production of nitrogen oxides in combustion and the manufacture and use of nitrogen fertilizers. As seen in Fig. 7, this adds fixed nitrogen to the inorganic soil reservoir from which it begins cycling through the ecosystem. In 1959 the use of fertilizers was responsible for the fixation of 3.5 Mt/yr of N. This had grown to 40 Mt(N)/yr by 1974 [67] and could reach 200 Mt(N)/yr by 2000. Biological fixation rates are uncertain, particularly for the oceans. The fixation rate due to the land biomass has been put at 44 Mt(N)/yr by Delwiche [66] and at 175 Mt(N)/yr by Hardy and Havelka [68]. The marine fixation rate has been estimated to be 10 Mt(N)/yr [61]. Thus if the lower biological fixation rates are correct, man's contribution to nitrogen fixation is already equal to the natural source, or, if the higher biological fixation rates are accepted, it will become so by the year 2000. The consequences of this potentially large anthropogenic source of

fixed nitrogen have been the subject of several papers [69] - [72].

Fig. 7 indicates four major reservoirs which contain the nitrogen cycling through the ecosystem: atmosphere, land, ocean, and biosphere. The nitrogen burden of these reservoirs is indicated in Mt in the small boxes, and fluxes between reservoirs are in Mt/yr and are indicated by the numbers labeling the arrows.

The atmosphere contains 4×10^9 Mt of nitrogen mostly as N_2, but there is a subreservoir (not shown explicitly in Fig. 7) of fixed atmospheric nitrogen which contains about 1 Mt(N) and consists primarily of NO_x and NH_3. The flow of nitrogen into this atmospheric subreservoir from the atmosphere results from lightning discharges (10-40 Mt/yr [64], [65]) and combustion (20.7 Mt/yr [67]). Our interest in perturbations to the nitrogen cycle center on the interactions of the biosphere with the other reservoirs and on the flow of nitrogen within the subreservoirs of the biosphere itself. The fixation flux, both natural and fertilizer, into the biosphere refers to processes of conversion of nitrogen from the biologically unutilizable N_2 to forms in which it can be used as nutrients, ammonium (NH_4^+), nitrate (NO_3^-), and nitrite (NO_2^-). The reverse process is denitrification, and the denitrification products may be either N_2 or N_2O.

Within the biosphere we have indicated three reservoirs: the living biomass, which contains about 7500 Mt (N); dead organic matter, called humus on land, which contains about 10^5 Mt (N), and the inorganic nitrogen pool, smallest of the three subreservoirs, containing about 5000 Mt (N). The flow of nitrogen into the biomass by natural fixation results from the activities of a multitude of micro-organisms, either free-living or living in association with plants, that are able to utilize the energy stored in the products of photosynthesis to break the N_2 bond and fix nitrogen into forms which may be used for the synthesis of protein, amino acids, and other organic compounds. This process essentially represents a flow of nitrogen directly from the atmosphere to the biomass subreservoir. Fertilizer, by contrast, adds fixed nitrogen to the inorganic soil reservoir which consists of soluble nitrate and nitrite. From the inorganic nitrogen pool the nitrogen may be assimilated by the biomass, which is the goal of fertilization, it may be immobilized, which means transferred to the organic nitrogen reservoir by bacterial assimilation or by sequestering of ammonia in clay particles; it may be lost by leaching from the soil, which may lead to problems such as the eutrophication of lakes; or it may be denitrified by denitrifying bacteria. The present debate on the impact of fertilizer nitrogen centers on the relative efficiencies and characteristic times for these loss processes.

The denitrification flux from the inorganic soil reservoir to the atmosphere is shown with a question mark in Fig. 7 since, although published estimates occur in the literature, these vary substantially (e.g., 43 Mt/yr [66], 210 Mt/yr [7]). Even if the flux is large it is possibly important only if a significant part of it consists of N_2O rather than N_2. Significance, in this context, must be judged relative to the magnitude of other sources and sinks of atmospheric N_2O and these, as noted earlier, are themselves a subject of debate. Finally, if the above parameters were firmly established, there is still the uncertainty of the effect of a given N_2O increase on stratospheric ozone.

Johnston [59] has recently published a review of this subject in which he sought to place limits on the values of the variables described above. The fraction of nitrogen fertilizer that is rapidly denitrified, β in Johnston's analysis, is highly uncertain. For a worst case analysis of ozone reduction β may be taken as 1. The fraction of N_2O in denitrified gases α is taken to be in the range $0.025 < \alpha < 0.4$ by Johnston based on his review of the available data, but as Johnston notes, these are estimates, not rigorous bounds. The importance of denitrified N_2O relative to other sources and sinks is discussed by Johnston in terms of the equivalent problem of the atmospheric residence time of N_2O. Large natural sources and sinks, which would minimize the impact of a given nitrous oxide flux in denitrified gases, correspond to shorter values of the residence time τ. From consideration of observed temporal and spatial variabilities of N_2O on the one hand (implying a short residence time and giving a lower bound to τ [57]) and the rate of stratospheric destruction on the other (giving the upper bound), the range of N_2O residence time given by Johnston is $5 \text{ yr} < \tau < 160 \text{ yr}$. Johnston seeks to place limits on the combinations of α and τ consistent with identified nitrogen fixation sources and concludes that for reasonable values of these variables the maximum ozone reduction due to added 100 Mt (N)/yr of fertilizer will lie between 0.4 and 12 percent and will occur within a few decades for 1 to 2 percent reduction or over hundreds of years for larger (~10-percent) reductions.

This analysis is based on estimates of the uncertainties involved in the parameters of the biospheric portion of the nitrogen cycle and does not discuss the possible impact of the uncertainties in the many chemical rate constants involved in the stratospheric modeling of ozone reductions [73], [74] (as Johnston clearly states). However, as we noted in our discussion of tropospheric ozone, the revision of a single chemical rate constant involved in a complex system can substantially alter the model conclusions. It is not possible to be even order-of-magnitude quantitative regarding the impact of nitrogenous fertilizers on stratospheric ozone and even Johnston's wide limits may be overly restrictive.

The nitrogen and carbon cycles are strongly coupled through the biosphere, but potential consequences of this fact have yet to be quantitatively explored. We have previously alluded to one aspect of this coupling in describing the constancy of the oceanic biomass in the face of increased CO_2 input to the oceans. This presumed constancy results from the fact that nitrogen, rather than carbon, is a limiting nutrient in the oceans. For example, the 5000 Mt/yr of carbon currently entering the atmosphere due to fossil fuel burning (Fig. 2) could be assimilated entirely by the oceanic biomass if an increased nitrogen input of 400 Mt/yr to the oceans occured assuming a C/N ratio of 12 for the ocean biomass. This is ten times the 1974 nitrogen fertilizer use and twice the projected use for the year 2000 [67]. The assimilation of this amount of nitrogen and carbon would require a yearly increase of 4 percent in the ocean biomass. There is yet insufficient data to determine the actual runoff of nitrate into the oceans, but the data discussed by Commoner [76] on nitrates in US river

systems show that this runoff has increased by a factor of 3-5 in the past 20 years.

Phosphorous, another limiting nut. .nt for the ocean biomass, is also being discharged into the oceans from land areas. The SCEP report [75] gives the 1968 US mean phosphorous content of runoff as 0.44 Mt.

In view of the increasing anthropogenic supply of limiting nutrients to the oceanic biomass it is conceivable that the increase in atmospheric CO_2 could be moderated in future years but the amounts of these nutrients required to reverse the effect are relatively large compared to projected usage. This problem deserves quantitative study with attention to the role of phosphorous as a limiting oceanic nutrient and to the fraction of fertilizer nitrogen which finds its way to the oceans.

VI. The Hydrogen Group

The most abundant member of this group is water vapor which is present in the atmosphere as a consequence of evaporation of liquid water at the earth's surface. Water vapor is chiefly responsible for the greenhouse effect through its strong absorption of infrared radiation, and it is an essential precursor for the OH radical produced through the reaction $O(^1D) + H_2O \longrightarrow 2OH$. The amount of H_2O in the atmosphere is highly variable and little influenced by human activities and its properties, therefore, will not be described in further detail.

Molecular hydrogen is the next most abundant member of this group, having a concentration of about 0.5 ppm. The prevailing view [77] is that H_2 is formed mainly by bacterial fermentation in soils and from the photodissociation of formaldehyde, one of the intermediate products in the photooxidation of methane, but the relative importance of these two paths is uncertain. The sink for H_2 is reaction with OH: $H_2 + OH \longrightarrow H + H_2O$.

The most important member of this group in influencing the concentrations of many trace tropospheric species is the hydroxyl radical OH. Reaction with OH is the dominant removal mechanism for constituents which are shielded from dissociation and are not removed via interaction with the hydrologic cycle. This is the case for many of the gases released into the atmosphere as a result of human activities, such as CO, NO_2, various halogenated methane compounds ($CHFCl_2$, CHF_2Cl, $CHCl_3$, CH_3Cl), halogenated ethylenes, and CH_3CCl_3. In many cases, reaction of the halogenated hydrocarbons with tropospheric OH removes a large potential source of stratospheric chlorine and thus reduces a potential harmful impact on the ozone layer. Recent measurements [33]-[35] show tropospheric mean OH concentrations on the order of 10^6 cm^{-3}. Reaction with OH is also the dominant removal mechanism for reduced gases of biological origin such as CH_4 and H_2S.

Hydroperoxy radicals are formed mainly by the reaction of CO and H_2 with OH. Because anthropogenic emissions of CO are now considered to represent a large fraction of the total atmospheric CO source, the potential exists for man's intervention in determining the background abundances of OH and HO_2 radicals in the troposphere. HO_2 radicals are not as reactive towards many species as are OH radicals and a shift in the odd hydrogen ($OH + HO_2$) balance could have consequences in determining the abundances of compounds mainly lost by reaction with OH, as was pointed out recently by Sze [25], Chameides *et al.* [26], and Hameed *et al.* [55].

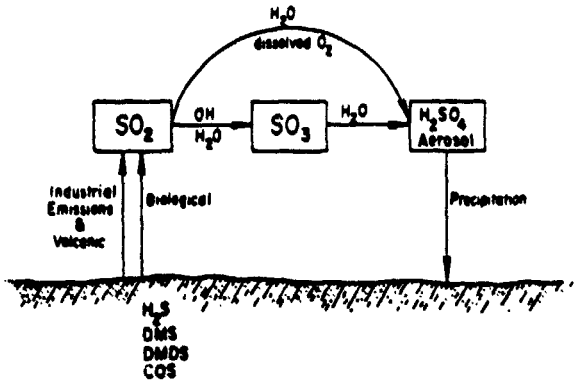


Fig. 8. Tropospheric sulfur chemistry over land. Biological sources emit the reduced sulfides H_2S, DMS (dimethylsulfide), DMDS (dimethyldisulfide), and COS (carbonyl sulfide). Industrial and volcanic sources emit SO_2 directly.

VII. The Sulfur Group

The abundances, sources and sinks, and photochemistry of atmospheric sulfur compounds are all poorly understood relative to oxygen, carbon, or nitrogen compounds. Various sulfur species are important components of air pollution in many parts of the world. The anthropogenic input to the sulfur cycle is clearly significant, as we shall see, and likely to increase in magnitude, but the consequences of these facts have been relatively little explored. The details of the sulfur cycle have been reviewed by Friend [78], Kellogg *et al.* [79], Robinson and Robbins [19], and Junge [8].

The importance of sulfur as a pollutant derives mainly from the properties of SO_2. This gas is toxic to plants and animals and is oxidized in the atmosphere to sulfuric acid. It thus acts as an aerosol precursor, may have an effect on cloud formation, and is therefore of potential climatological significance. The H_2SO_4 may lower the pH of rainwater with deleterious consequences. Toxicity of SO_2 and acid rain are likely to be confined to local and regional scale effects due to the short (~1-week) residence time of atmospheric sulfur, but changes in aerosol or cloud properties are global scale phenomena.

The chemistry of the atmospheric portion of the sulfur cycle over land is shown in Fig. 8. Volcanic activity is a modest source of natural SO_2 (2 Mt (S)/yr) compared with industrial output (65 Mt (S)/yr). The largest natural source of atmospheric sulfur is provided by decay of organic material in the biosphere (58 Mt (S)/yr over land). This input is in the form of reduced sulfur such as hydrogen sulfide and organic sulfides (dimethyl sulfide, DMS, dimethyl disulfide, DMDS) and is rapidly oxidized to SO_2 by reactions with atmospheric trace constituents [80] The oxidation mechanism is as yet unclear. The lifetime of reduced sulfur in the atmosphere is of the order of hours, but reaction of H_2S with O_3 results in a lifetime of a month or more. Reaction of H_2S with OH is faster but still results in a lifetime of a few days. It may be that oxidation of reduced sulfur in the atmosphere

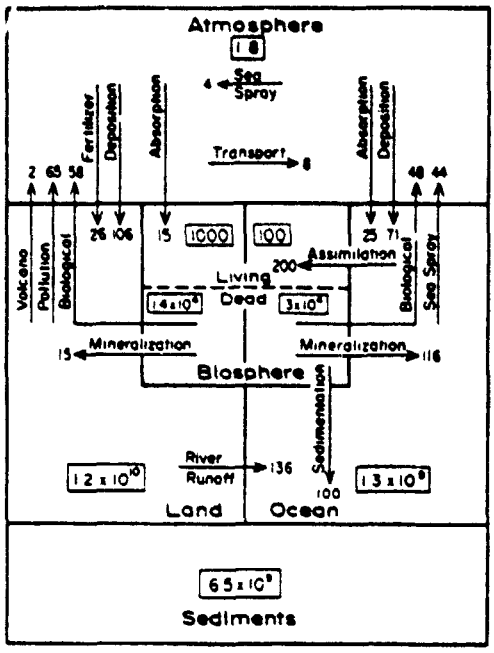


Fig. 9. The Sulfur cycle. The number of Mt of sulfur in the indicated reservoirs are shown in boxes. The fluxes between reservoirs are shown by the arrows and given in Mt (S)/yr. The physical or biological process giving rise to these fluxes is stated next to the arrow.

results from reaction of organic sulfides with NO_x, but this mechanism is uncertain.

Sulfur dioxide is itself oxidized in the atmosphere with a lifetime of hours. This may occur by gas phase reaction with OH and HO_2 radicals to form SO_3, which is then rapidly absorbed in water vapor to form sulfuric acid, or the SO_2 may undergo heterogeneous removal by reaction with dissolved O_2 in water droplets. The relative importance of these mechanisms is uncertain, but it is difficult to account for the short SO_2 lifetime solely by gas phase chemistry.

A schematic diagram of the sulfur cycle is shown in Fig. 9. The magnitudes of the sulfur reservoirs and fluxes are based on the model of Friend [78] but have been adjusted where necessary to maintain the observed N/S ratio of 7 [81] The large impact which human activity has on the sulfur budget is obvious. The 65 Mt/yr of sulfur emitted to the atmosphere has a substantial influence on pollution on the urban and regional scale, but whether there are global effects associated with this or with potentially larger SO_2 inputs is a largely unexplored problem. The 26 Mt/yr of sulfur added to the soil in fertilizers is a significant and growing fraction of the total soil input of 147 Mt/yr. Possible ecological consequences of this addition have not been addressed.

VIII. The Halogen Group

The concentration of halogen compounds in the atmosphere is much less than that of the other groups thus far discussed, being generally of the order of ppb or less. Recent research interest in this group derives from the fact that the anthropogenic contribution to atmospheric halogens is a clearly significant source and that these species are believed to efficiently reduce stratospheric ozone and in some cases to possess radiative properties of possible climatic significance [82].

The most abundant naturally occurring halogens are the marine organic halogens CH_3Cl, CH_3Br, and CH_3I [36] which are present in abundances of about 780 ppt, 10 ppt, and 1.2 ppt, respectively.

Anthropogenic contributions to the halogen budget consist primarily of the chlorofluoromethanes CCl_3F and CCl_2F_2 and carbon tetrachloride CCl_4 [83]. In the early 1970's the tropospheric concentrations of CCl_3F, CCl_2F_2, and CCl_4 were about 90 ppt, 100 ppt, and 120 ppt, respectively [84].

Most of the concern over atmospheric halogens centers on the ability of the long-lived industrial compounds CCl_3F and CCl_2F_2 to diffuse into the stratosphere where they will dissociate and release chlorine, leading to a postulated reduction of ozone. A review of the extensive literature on this subject is not within the scope of this paper (see [85], [86]).

The accumulation of chlorofluorocarbons and CCl_4 in the troposphere could have climatic consequences, however, as pointed out by Ramanathan [82], who states that a surface temperature increase of 0.8 K could occur if the concentration of CCl_3F and CCl_2F_2 were each increased to 2 ppb, about a 20-fold increase over present levels. Wang *et al.* [56] derive a 0.54 K maximum surface temperature increase for the same assumed increase in CCl_3F and CCl_2F_2 concentrations. The difference in these two results is to some extent due to the fact that Wang *et al.* use more recently measured band intensities for the ν_1 (9.13 μm) and ν_8 (10.93 μm) bands of CCl_2F_2 and the ν_1 (9.22 μm) and ν_4 (11.82 μm) bands of CCl_3F which are roughly half the older values used by Ramanathan.

IX. Summary

The fact of man's intervention in the geochemical cycles which control the composition of the atmosphere has only recently been appreciated. Attempts to predict the consequences of such interventions now occupy the efforts of a growing number of researchers in many fields. These efforts are necessarily of a broadly interdisciplinary character involving biologists, chemists, geologists, oceanographers, meteorologists, aeronomers, and others seeking answers to the many questions of environmental concern which have been raised by man's activities.

Interest in the oxygen group from the point of view of its influence on pollution and climate centers on the role of ozone and atomic oxygen in tropospheric photochemistry. Ozone is the precursor of all global-scale tropospheric photochemistry and understanding of the processes which produce and destroy ozone in the troposphere is thus of fundamental importance as an element in the understanding of the atmospheric phase of other geochemical cycles. This problem is currently an active area of research, but the major questions regarding the relative importance of transport and photochemistry in establishing the ozone distribution remain unanswered.

Man's intervention in the carbon cycle results chiefly from the burning of fossil fuels and is mainly manifested in the increasing atmospheric concentration of carbon dioxide. The general problem is to predict future atmospheric levels of CO_2 and the climatic consequences of such levels. The major

issues involved in predicting these levels presently include the role of deforestation as a CO_2 source and the present and future effectiveness of the oceans as a CO_2 sink. The major difficulties in predicting the climatic consequences of increased CO_2 abundances are in properly including climate feedback mechanisms, such as albedo and cloud cover changes and ocean coupling, in mathematical models.

A perturbation to the nitrogen cycle is occurring as a result of the increasing use of fertilizers in agriculture. The annual industrially fixed nitrogen resulting from this use is either already or by the end of the century will be equal to the amount of nitrogen fixed naturally. It has been suggested that one consequence of this intervention in the nitrogen cycle will be an increase in the abundance of nitrous oxide as the denitrification rate adjusts to the increased fixation rate. Additional N_2O could result in a decrease in the amount of stratospheric ozone and might also make a contribution to the greenhouse effect. Uncertainties abound in this problem. Among those that have been discussed in the literature are the atmospheric residence time of N_2O (5-160 yrs), the fraction of denitrified fertilizer nitrogen that enters the atmosphere as N_2O (0.025-0.4), the fraction of fertilizer nitrogen that is rapidly denitrified (highly uncertain), and the effect of a given N_2O increase on stratospheric ozone and the greenhouse effect (perhaps a factor of 2 uncertainty).

The hydrogen group is chemically active in the atmosphere and is thus subject to perturbations. Changes in the abundances of the components of this group would not (excepting water vapor) have a direct influence on pollution or climate. The extremely active hydroxyl radical, however, reacts strongly with every other major atmospheric group and changes in its mean abundance could have far-reaching consequences.

Man is also intervening in the sulfur cycle, but global-scale consequences of this influence have yet to be identified. Most of the SO_2 that enters the atmosphere is of industrial origin and this can potentially alter such things as the atmospheric aerosol loading the acidity of rainfall over at least mesoscale regions. The 26 Mt/yr of sulfur added to the soil in fertilizers is relatively large compared to natural sources. If substantially increased coal burning occurs in the future in response to shortages of alternate fuels the magnitude of the industrial sulfur source will increase still further. It is essential that some attempts to understand the potential impact of this fact be made in future research.

The halogens are relatively inactive chemically in the troposphere and recent research interest has centered on the reduction of stratospheric ozone which might be a consequence of their release into the atmosphere. Some of the components of this group do, however, have absorption bands in the thermal IR and it has been suggested that their continued buildup in the troposphere might contribute more than half a degree Kelvin to the greenhouse effect.

It must be borne in mind that the elemental cycles of oxygen, carbon, nitrogen, hydrogen, and sulfur which have been discussed individually are strongly coupled to one another through the biosphere and atmosphere. The carbon and oxygen cycles are coupled over time scales up to $\sim 10^3$ years by the biological processes of photosynthesis, respiration and decay, and over longer time scales by geochemical interactions. The carbon, nitrogen, and sulfur cycles are coupled by the nutrient needs of the biosphere and to a lesser extent by chemical interactions in the atmosphere. The increasing abundance of carbon dioxide resulting from human intervention in the carbon cycle may not in itself permit an expansion of the biomass since other nutrients such as nitrogen and phosphorous are the limiting factors. On a small ecological scale the eutrophication of lakes and streams provide examples of the extent to which natural cycles may be perturbed when limiting nutrients are supplied by man. With the possible exception of the interactions of the carbon-oxygen cycle, the interactions between the elemental cycles have not yet been explored in adequate quantitative detail. It is likely that as understanding increases, interrelationships among the various geochemical cycles will emerge and the interdisciplinary nature of this research will be enhanced.

References

[1] H. Levy, "Photochemistry of the lower troposphere," *Planet. Space Sci.*, vol. 20, pp. 919-935, 1972.

[2] —, "Photochemistry of minor constituents in the troposphere," *Planet. Space Sci.*, vol. 21, pp. 575-591, 1973.

[3] H. Heicklin, *Atmospheric Chemistry*. New York: Academic Press, 1976.

[4] J. C. G. Walker, "Stability of atmospheric oxygen," *Amer. J. Sci.*, vol. 274, pp. 193-214, 1974.

[5] R. M. Garrels, A. Lerman, and F. T. MacKenzie, "Controls of atmospheric O_2 and CO_2: Past, present, and future," *American Scientist*, vol. 64, pp. 306-315, May-June, 1976.

[6] C. E. Junge, M. Schidlowski, R. Eichmann, and H. Pietrek, "Model calculations for the terrestrial carbon cycle: Carbon isotopes geochemistry and evolution of photosynthetic oxygen," *J. Geophys. Res.*, vol. 80, pp. 4542-4552, Nov. 1975.

[7] M. B. McElroy, "Chemical processes in the solar system: A kinetic perspective," in *MTP International Review of Science*, D. Herschbach, Ed. 1976.

[8] C. E. Junge, *Air Chemistry and Radioactivity*. New York: Academic Press, 1963.

[9] W. L. Chameides and J. C. G. Walker, "A photochemical theory of tropospheric ozone," *J. Geophys. Res.*, vol. 78, pp. 8751-8760, Dec. 1973.

[10] W. L. Chameides and J. C. G. Walker, "A time-dependent photochemical model for ozone near the ground," *J. Geophys. Res.*, vol. 81, pp. 413-420, Jan. 1976.

[11] P. A. Leighton, *Photochemistry of Air Pollution*. New York: Academic Press, 1961.

[12] P. J. Crutzen, "Gas phase nitrogen and methane chemistry in the atmosphere," *Physics and Chemistry of the Upper Atmosphere*, B. McCormac, Ed. Dordrecht, Netherlands: D. Reidel, 1973.

[13] J. Fishman and P. J. Crutzen, "A numerical investigation of tropospheric photochemistry using a one-dimensional model," in *The Non-Urban Tropospheric Composition*, A joint symposium of the American Geophysical Union and the American Meteorological Society, Hollywood, FL, Nov. 10-12, 1976.

[14] P. Fabian, "Comments on a 'Photochemical theory of tropospheric ozone' by W. Chameides and J. C. G. Walker," *J. Geophys. Res.*, vol. 79, pp. 4124-4125, 1974.

[15] P. Fabian and P. G. Pruchniewicz, "Meridional distribution of ozone in the troposphere and its seasonal variations," *J. Geophys. Res.*, vol. 82, pp. 2063-2073, 1977.

[16] W. L. Chameides and D. H. Stedman, "Tropospheric ozone: Coupling transport and photochemistry," *J. Geophys. Res.*, vol. 82, pp. 1787-1794, Apr. 1977.

[17] R. W. Stewart, S. Hameed, and J. P. Pinto, "Photochemistry of tropospheric ozone," *J. Geophys. Res.*, vol. 82, pp. 3134-3140, July 1977.

[18] R. F. Hampson and D. Garvin, "Chemical kinetic and photochemical data for modelling atmospheric chemistry," *NBS Technical Note 866*, Nat. Bur. Stand., Washington, DC, 1975.

[19] E. Robinson and R. C. Robbins, "Sources, abundance, and fate of gaseous atmospheric pollutants," *Stanford Research Institute Final Rept. Proj. PR-6755*, 1968.
[20] J. F. Noxon, "Nitrogen dioxide in the stratosphere and troposphere measured by ground-based absorption spectroscopy," *Science*, vol. 185, pp. 547-549, Aug. 1975.
[21] C. J. Howard and K. M. Evenson, "Laser magnetic resonance study of HO_2 chemistry," *EOS, Trans. Amer. Geophys. Union*, *58*, p. 464, June 1977.
[22] E. F. Danielsen and V. A. Mohnen, "Project dustorm report: Ozone measurements and meteorological analysis of tropopause folding," in *The Non-Urban Tropospheric Composition*, A joint symposium of the American Geophysical Union and The American Meteorological Society, Hollywood, FL, Nov. 10-12, 1976.
[23] K. Westberg, N. Cohen, and K. W. Wilson, "Carbon monoxide: Its role in photochemical smog formation," *Science*, vol. 171, pp. 1013-1015, Mar. 1975.
[24] W. A. Glasson, "Effect of carbon monoxide on atmospheric photooxidation of nitric oxide-hydrocarbon mixtures," *Environ. Sci. Tech.*, vol. 9, 343-347, Apr. 1975.
[25] N. D. Sze, "Anthropogenic CO emissions: Implications for the atmospheric CO-OH-CH_4 cycle," *Science*, vol. 195, pp. 673-675, Feb. 1977.
[26] W. L. Chameides, S. C. Liu, and R. J. Cicerone, "Possible variations in atmospheric methane, *J. Geophys. Res.*, vol. 82, pp. 1795-1798, Apr. 1977.
[27] B. Bolin, "The Carbon Cycle," *Scientific American*, vol. 223, pp. 124-132, Sept. 1970.
[28] J. Cramer and A. L. Myers, "Rate of increase of atmospheric carbon dioxide," *Atmos. Environ.*, vol. 6, pp. 563-573, 1972.
[29] M. I. Hoffert, "Global distribution of atmospheric carbon dioxide in the fossil-fuel era: A projection," *Atmos. Environ.*, vol. 8, pp. 1225-1249, 1974.
[30] C. D. Keeling, "The carbon dioxide cycle: Reservoir models to depict the exchange of atmospheric carbon dioxide with the ocean and land plants," *Chemistry of the Lower Atmosphere*, S. I. Rasool, Ed. New York: Plenum Press, 1973.
[31] C. F. Baes, Jr., H. E. Goeller, J. S. Olson, and R. M. Rotty, "Carbon dioxide and climate: The uncontrolled experiment," *American Scientist*, vol. 65, pp. 310-320, May-June 1977.
[32] D. H. Ehhalt, "The atmospheric cycle of methane," *Tellus*, vol. 26, pp. 58-70, 1974.
[33] C. C. Wang, L. I. Davis, Jr., C. H. Wu, S. Japar, H. Niki, and B. Weinstock, "Hydroxyl radical concentrations measured in ambient air," *Science*, vol. 189, pp. 797-800, Sept. 1975.
[34] D. D. Davis, W. Meaps, and T. McGee, "Direct measurements of natural tropospheric levels of OH via an aircraft borne tunable dye laser," *Geophys. Res. Lett.*, vol. 3, pp. 331-333, June 1976.
[35] D. Perner, D. H. Ehhalt, H. W. Patz, V. Platt, E. P. Roth, and A. Voly, "OH-radicals in the lower troposphere," *Geophys. Res. Lett.*, vol. 3, pp. 466-468, Aug. 1976.
[36] H. B. Singh, "Atmospheric halocarbons: Evidence in favor of reduced average hydroxyl radical concentration in the troposphere," *Geophys. Res. Lett.*, vol. 4, pp. 101-104, Mar. 1977.
[37] P. Crutzen, "Photochemical reactions initiated by and influencing ozone in unpolluted tropospheric air," *Tellus*, vol. 26, pp. 47-57, 1974.
[38] W. Seiler, "The cycle of atmospheric CO," *Tellus*, vol. 26, pp. 116-135, 1974.
[39] S. C. Wofsy, "Interactions of CH_4 and CO in the earth's atmosphere," *Annual Review of Earth and Planetary Sciences*, F. Donath, Ed. Vol. 4, pp. 441-469, 1976.
[40] P. Warneck, "On the role of OH and HO_2 radicals in the troposphere," *Tellus*, vol. 26, pp. 39-46, 1974.
[41] R. E. Newell, "One-dimensional models: A critical comment, and their application to carbon monoxide," *J. Geophys. Res.*, 9, pp. 1449-1450, Mar. 1977.
[42] W. Bischof and B. Bolin, "Space and time variations of the CO_2 content of the troposphere and lower stratosphere," *Tellus*, vol. 18, pp. 155-159, 1966.
[43] B. Bolin and W Bischof, "Variation of the carbon dioxide content of the atmosphere in the northern hemisphere," *Tellus*, vol. 22, pp 431-442, 1970
[44] L. Machta, "The role of the oceans and biosphere in the carbon cycle," presented at 20th Nobel Symp., Changing of the Oceans, Gothenburg, Sweden, Aug. 16-20, 1971.
[45] C. D. Keeling and Bacastow, "Impact of industrial gases on climate," *National Academy of Sciences Report*, in press, 1977.
[46] J. A. S. Adams, M. S. M. Mantovani, and L. L. Lundell, "Wood versus fossil fuel as a source of excess carbon dioxide in the atmosphere: A preliminary report," *Science*, vol. 196, pp. 54-56, Apr. 1977.
[47] B. Bolin, "Change of land biota and their importance for the carbon cycle," *Science*, vol. 196, pp. 613-615, May 1977.
[48] W. Broecker, "The climate consequences of a coal economy," presented at *Spring Meeting of AGU*, Washington, DC, June 6-10, 1977.
[49] S. I. Rasool and S. H. Schneider, "Atmospheric carbon dioxide and aerosols: Effects of large increases on global climate," *Science*, vol. 173, pp. 138-141, 1971.
[50] S. Manabe and R. T. Wetherald, "The effects of doubling the CO_2 concentration on the climate of a general circulation model," *J. Atmos. Sci.*, vol. 32, pp. 3-15, Jan. 1975.
[51] F. Moller, "On the influence of changes in CO_2 concentration in air on the radiative balance of the earth's surface and on the climate," *J. Geophys. Res.*, vol. 68, pp. 3877-3886, 1963.
[52] S. Manabe, "Estimate of future changes of climate due to increase of carbon dioxide concentration in the air," *Man's Impact on Climate*, W. H. Matthews, W. W. Kellogg, and G. D. Robinson, Eds. MIT Press, 1971, p. 256.
[53] S. H. Schneider, "On the carbon dioxide-Climate confusion," *J. Atmos. Sci.*, vol. 32, pp. 2060-2066, Nov. 1975.
[54] T. Augustsson and V. Ramanathan, "A radiative-convective model study of the CO_2 climate problem," *J. Atmos. Sci.*, vol. 34, pp. 448-451, Mar. 1977.
[55] S. Hameed, J. Pinto, and R. W. Stewart, "Carbon monoxide emissions: Modification of the CO-CH_4 perturbation by NO_x," *EOS, Trans. Am. Geophys. Union*, vol. 58, p. 396, June 1977.
[56] W. C. Wang, Y. L. Yung, A. A. Lacis, T. Mo, and J. E. Hansen, "Greenhouse effects due to man-made perturbation of trace gases," *Science*, vol. 194, pp. 685-690, Nov. 1976.
[57] C. E. Junge, "Residence time and variability of tropospheric trace gases," *Tellus*, vol. 26, pp. 477-488, 1974.
[58] K. Schutz, C. E. Junge, R. Beck, and B. Albrecht, "Studies of Atmospheric N_2O," *J. Geophys. Res.*, vol. 75, pp. 2230-2246, 1970.
[59] H. S. Johnston, "Analysis of the independent variables in the perturbation of stratospheric ozone by nitrogen fertilizers," *J. Geophys. Res.*, vol. 82, pp. 1767-1772, Apr. 1977.
[60] J. Hahn, "The North Atlantic ocean as a source of atmospheric N_2O," *Tellus*, vol. 26, pp. 160-168, 1974.
[61] M. B. McElroy, J. W. Elkins, S. C. Wofsy, and Y. L. Yung, "Sources and sinks for atmospheric N_2O," *Rev. Geophys. Space Phys.*, vol. 14, pp. 143-150, 1976.
[62] H. W. Georgii and W. J. Müller, "On the distribution of ammonia in the middle and lower troposphere," *Tellus*, vol. 26, pp. 180-184, 1974.
[63] G. A. Dawson, "Atmospheric ammonia from undisturbed land," *J. Geophys. Res.*, vol. 82, pp. 3125-3133, July 1977.
[64] J. F. Noxon, "Atmospheric nitrogen fixation by lightning," *Geophys. Res. Lett.*, vol. 3, pp. 463-465, 1976.
[65] W. L. Chameides, D. H. Stedman, R. R. Dickerson, D. W. Busch, and R. J. Cicerone, "NO_x production in lighting," *J. Atmos. Sci.*, vol. 34, pp. 143-149, Jan. 1977.
[66] C. C. Delwiche, "The nitrogen cycle," *Scientific American*, vol. 223, pp. 136-146, Sept. 1970.
[67] CAST (Council for Agricultural Science and Technology), "Effect of increased nitrogen fixation on stratospheric ozone," Report No. 53, Jan. 19, 1976.
[68] R. W. F. Hardy and V. D. Havelka, "Nitrogen fixation research: A key to world food," *Science*, pp. 633-643, May 1975.
[69] S. C. Liu, R. J. Cicerone, T. M. Donahue, and W. L. Chameides, "Limitation of fertilizer induced ozone reduction by the long lifetime of the reservoir of fixed nitrogen," *Geophys. Res. Lett.*, vol. 3, pp. 157-160, Mar. 1976.
[70] P. J. Crutzen, "Upper limits on atmospheric ozone reduction following increased application of fixed nitrogen to the soil," *Geophys. Res. Lett.*, vol. 3, pp. 169-172, Mar. 1976.
[71] N. D. Sze and H. Rice, "Nitrogen cycle factors contributing to N_2O production from fertilizers," *Geophys. Res. Lett.*, vol. 3, pp. 343-346, June 1976.
[72] S. C. Liu, R. J. Cicerone, T. M. Donahue, and W. L. Chameides, "Sources and sinks of atmospheric N_2O and the possible ozone reduction due to industrial fixed nitrogen fertilizers," *Tellus*,

29, pp. 251–263, 1977.
[73] W. H. Deuwer, D. J. Wuebbles, H. W. Elsaesser, and J. S. Chang, "NO_x catalytic ozone destruction: Sensitivity to rate coefficients," *J. Geophys. Res.*, *82*, pp. 935–942, 1977.
[74] H. S. Johnston and H. Nelson, "Comment on 'NO_x catalytic ozone destruction: Sensitivity to rate coefficients' by W. H. Duewer, D. J. Wuebbles, H. W. Elsaesser, and J. S. Chang," *J. Geophys. Res.*, *82*, pp 2593–2598, 1977.
[75] SCEP, *Study of Critical Environmental Problems*. Cambridge, MA: MIT Press, 1970.
[76] B. Commoner, "Threats to the integrity of the nitrogen cycle: Nitrogen compounds in soil, water, atmosphere, and precipitation," *The Changing Global Environment*, S. F. Singer, Ed. Boston: D. Reidel, 1975.
[77] V. Schmidt, "Molecular hydrogen in the atmosphere," *Tellus*, vol. 74, pp. 78–90, 1974.
[78] J. P. Friend, "The global sulfur cycle," *Chemistry of the Lower Atmosphere*, S. I. Rasool, Ed. New York: Plenum Press, 1973, pp. 177–201.
[79] W. W. Kellogg, R. D. Cadle, E. R. Allen, A. L., Lazrus, and E. A. Martell, "The sulfur cycle," *Science*, *175*, pp. 587–596, Feb. 1972.
[80] R. A. Rasmussen, "Emission of biogenic hydrogen sulfide," *Tellus*, vol. 26, pp. 254–260, 1974
[81] G. Anderson, "Sulfur in soil organic substances," *Soil Components, vol. 1, Organic Components*, J. E. Gieseking, Ed. New York: Springer-Verlag, 1975.
[82] V. Ramanathan, "Greenhouse effect due to chlorofluorocarbons: Climatic implications," *Science*, vol. 190, pp. 50–52, Oct. 1975.
[83] F. S. Rowland and M. J. Molina, "Chlorofluormethanes in the environment," *Rev. Geophys. and Space Sci.*, vol. 13, pp. 1–35, Feb. 1975.
[84] J. E. Lovelock, "Atmospheric halocarbons and stratospheric ozone," *Nature*, vol. 252, pp. 292–294, Nov. 1974.
[85] *Halocarbons: Effects on Stratospheric Ozone*, National Academy of Sciences, Washington, DC, 1976.
[86] A. J. Grobecker, S. C. Coroniti, and R. H. Cannon, Jr., "The effects of stratospheric pollution by aircraft, findings of the climatic impact assessment program," *Report DOT-TST-75-50*, Dept. of Transp., Washington, DC, Mar. 1975.
[87] H. Carver, "Some problems of atmospheric chemistry," *Comp. Meteorology*, pp. 1126–1138.
[88] E. Robinson, R. A. Rasmussen, H. H. Westberg, and M. W. Holdren, "Non urban non-methane low molecular weight hydrocarbon concentrations related to air mass identification," *J. Geophys. Res.*, vol. 78, pp. 5345–5351, 1973.
[89] C. B. Farmer, B. F. Ryer, and R. H. Norton, "Spectroscopic detection and vertical distribution of HCl in the troposphere and stratosphere," *Geophys. Res. Lett.*, vol. 3, pp. 13–16, Jan. 1976.
[90] C. Junge, "The cycle of atmospheric gases–natural and man made," *Quart. J. Roy. Met. Soc.*, vol. 98, pp. 711–729, 1972.

THE ASTROPHYSICAL JOURNAL, 227:943–954, 1979 February 1

SOME SIMPLE PROPERTIES OF STELLAR PULSATION MODES

CHARLES L. WOLFF
NASA-Goddard Space Flight Center
Received 1978 May 15; accepted 1978 August 9

ABSTRACT

Except for the lowest harmonics, small-amplitude stellar pulsation modes possess many simple properties whose evaluation does not require numerical integration of the fourth-order equations of motion. All antinodes tend to have the same total kinetic energy except for those lying near physical or geometric boundaries. However, when kinetic energy per unit volume is considered, order-of-magnitude enhancements are seen in antinodes lying near the center of the star, and factor-of-2 enhancements occur near the polar axis. The nodes are distributed very regularly along the radius. They follow an exponential law in g-regions, and their separation is proportional to the sound travel time in p-regions. A simple graphical procedure is described for surveying the oscillation frequencies of a new stellar model. A precise condition is derived giving the division of energy between radial and angular motion. Another condition gives the fractional contribution to the velocity field of its two sources, the divergence and the curl. Certain simplifying results of weak coupling among the linear modes are briefly described.

Subject heading: stars: pulsation

I. INTRODUCTION

It is still possible to discover simple properties of stellar pulsation modes because the subject has developed along a unique course emphasizing detailed numerical integration of the lowest harmonic modes. While this course has led to successful interpretations of giant classical variables, there has been growing recognition in this decade that many stars are multiperiodic, and therefore require a broader outlook. White dwarf variables are known with many periods in their light curves; an example would be HL Tau 76, as measured by Fitch (1973). The β Canis Majoris stars also appear to be multiperiodic (see Ledoux 1974 for a brief review). Measurements by Deubner (1975) and Rhodes, Ulrich, and Simon (1977) prove that a great many modes are active in the Sun with periods from 3 to 10 minutes. Excitation of these global modes in the Sun was predicted by Wolff (1972*a*, *b*), and part of that work has been redone much more completely by Ando and Osaki (1975). Numerous oscillations of the Sun at periods up to an hour were detected by Hill and Stebbins (1975) and by Hill, Stebbins, and Brown (1976) in solar diameter measurements. Others did not confirm this, but Hill (1977) has made a complete analysis of reports from various observers and finds the most likely interpretation to be that long-period oscillations are active in the Sun.

Over the last half century, until recently, studies of variable stars were pursued under four strong biases: (1) Spherically symmetric motion was usually assumed for simplicity, although work before Eddington often discussed angular modes associated with tidal distortions (Ledoux and Walraven 1958). Relatively few authors, such as Cowling and groups working with Ledoux and Chandrasekhar, gave detailed consideration to nonradial motion. (2) The fundamental mode of oscillation and the several lowest harmonics had been studied almost exclusively. (3) Observers expected to find one period of oscillation or a few, at most. Many additional periodicities near the noise level could have easily passed undiscovered in the presence of this bias. (4) Coupling among the modes has been almost entirely ignored except for the elementary possibility of a pair of low harmonics whose oscillation periods happen to be almost the same. Dramatically different results can occur for higher-order modes under nonlinear coupling. For example, Wolff (1974*a*) imposed a nonlinear energy-maximizing condition to couple the modes. This led to motion concentrated into a few small solid angles centered on the star, leaving most of the 4π steradians participating only weakly in the oscillation. In this paper, we will consider the full spectrum of small-amplitude pulsation modes in a nonrotating star (the spheroidal modes). Several properties applying to all the modes are given in § II. Properties applying to all but the well studied low harmonics are collected in § III. The unique nature of the central and polar regions is discussed in § IV, and three forms of mode coupling are discussed in § V.

II. GENERAL RESULTS

a) Linearized Equations in Compact Form

For a star whose time-averaged state is at rest and in hydrostatic equilibrium, adiabatic pulsations of small amplitude are governed by the three linearized, fluid equations conserving mass, momentum, and energy,

$$0 = \rho + \rho_0 \nabla \cdot \xi + \xi \cdot \nabla \rho_0 , \tag{1}$$

$$0 = -\lambda \rho_0 \xi + \nabla p + \rho \nabla W_0 + \rho_0 \nabla W , \tag{2}$$

$$0 = p + \Gamma_1 p_0 \nabla \cdot \xi + \xi \cdot \nabla p_0 , \tag{3}$$

and by Poisson's equation, $\nabla^2(W_0 + W) = 4\pi G \times (\rho_0 + \rho)$. All time dependence is in a factor, $\exp(i2\pi\nu t)$, which has already been separated and canceled from the above system. In these equations, p_0, ρ_0, and W_0 represent the time-averaged pressure, density, and gravitational potential. The same symbols without subscripts and multiplied by the time dependence represent the Eulerian perturbations due to the oscillations. The displacement ξ is always used instead of the velocity $v = \partial\xi/\partial t$ because the displacement is in time-phase with the other pulsation variables. Finally, G is the gravitational constant and Γ_1 is the appropriate adiabatic coefficient. The graphs in this paper will show the true frequency ν of the oscillation. This or its inverse is what observers almost universally plot, and it is the basic physical quantity. But the square of the angular frequency appears often and it will be understood that

$$\lambda_l = (2\pi\nu_l)^2 , \tag{4}$$

for any frequency defined herein. Ledoux (1974) and Cox (1976) provide recent reviews of the solutions to this system of equations, and Hill (1977) discusses solutions in surface layers. For the Sun alone, ~100 oscillation frequencies have been computed, and Hill gives a graph of their distribution in frequency and angular harmonic number.

A linear oscillation mode depends on just one of the spherical harmonic functions, $Y_l^m(\theta, \phi)$, which can be separated and canceled from the above system. The resulting radial equations are greatly simplified by choosing a pair of natural variables,

$$P = \lambda^{1/2} E D^{-1} \xi_r , \tag{5}$$

$$Q = \lambda^{1/2} E r \xi_\theta , \tag{6}$$

where E is the adiabatic invariant $\rho_0 p_0^{-1/\Gamma_1}$ and $D = E^2 \rho_0^{-1} r^{-2}$. The variables are proportional to the radial and angular displacement functions, ξ_r and ξ_θ, defined by

$$\xi = \xi_r e_r Y_l^m + \xi_\theta r \nabla Y_l^m , \tag{7}$$

where e_r is a unit vector in the radial direction. This form of the displacement gives all the spheroidal modes but excludes the toroidal modes (Aizenman and Smeyers 1977). The latter become important in rapidly rotating stars. Ledoux and Walraven (1958, § 79) showed how variables similar to P and Q compress the second order equations. Our variables will similarly shorten most scalar expressions arising in the fourth-order system. The ratio, Q/E, is the amplitude of the angular momentum per unit mass caused by the oscillation. The product, PE, is the amplitude of radial momentum per unit solid angle and unit radial distance. The word "amplitude" could be removed if PE and Q/E were multiplied by the sinusoidal time dependence and by their respective angular dependences, Y and $e_r \times \nabla Y$. In the natural variables, equations (1)–(3) collapse to

$$P' + (\lambda - \lambda_l)\frac{Q}{Dc^2} = w_1 , \tag{8}$$

$$Q' - (\lambda - \lambda_0)\frac{DP}{\lambda} = w_2 , \tag{9}$$

where primes will hereafter stand for d/dr, terms containing derivatives of Γ_1 have been ignored, and $c^2 = \Gamma_1 p_0 \rho_0^{-1}$, the square of the sound speed. The right-hand members are proportional to the gravity perturbation and are usually negligible except at the center of the star. Their exact values are $w_1 = \lambda^{1/2} E D^{-1} c^{-2} W$ and $w_2 = \lambda^{-1/2} E' W$. The compact form of (8) and (9) has the additional virtue of clearly displaying the squares of the two frequencies controlling the radial distribution of kinetic energy in a global oscillation mode,

$$\lambda_0 = -W_0' E'/E , \tag{10}$$

$$\lambda_l = l(l+1)c^2/r^2 . \tag{11}$$

The frequencies are associated with regular singularities in the equations of motion (Ledoux 1974) and will be discussed and plotted in § III*a*.

b) Curl- and Divergence-Dominated Motion

The flow pattern in a particular oscillation mode is quite often mainly a swirling motion with little compression, or it is just the opposite—a mostly irrotational motion with significant compression. The relative importance of each type is easily derived. First, note that the last term in each of equations (1) and (3) depends only on ξ_r because the equilibrium state is spherically symmetric. Eliminating ξ_r from these two gives

$$\nabla \cdot \xi = \frac{E}{E' c^2 \rho_0^2}(p_0' \rho - \rho_0' p) . \tag{12}$$

Dividing (2) by $\lambda\rho_0$ and taking the curl gives

$$\nabla \times \xi = \frac{1}{\lambda \rho_0^2}(\nabla p_0 \times \nabla \rho - \nabla \rho_0 \times \nabla p) , \tag{13}$$

where use was made of the hydrostatic condition, $\rho_0 \nabla W_0 = -\nabla p_0$. It is known that the angular dependence of both ρ and p is Y_l^m for a given mode. Substituting this into (12) and (13) gives a result which can be written as the product of a radial and an angular function,

$$\nabla \cdot \xi = A_c(r)[Y_l^m] \tag{14}$$

$$\nabla \times \xi = A_c(r)[(l^*)^{-1} r \times \nabla Y_l^m] . \tag{15}$$

Using the factor $l^* = [l(l+1)]^{1/2}$ ensures that the absolute value squared of each quantity in brackets will integrate to one over the unit sphere. Equations (14) and (15) define a divergence amplitude, A_c, and

a curl amplitude, A_c. Forming the quotient of (12) and (13) and then introducing these definitions give the very simple result,

$$\frac{-A_d}{A_c} = \frac{\lambda}{\lambda_{cd}} = \left(\frac{\nu}{\nu_{cd}}\right)^2, \tag{16}$$

where $\lambda_{cd} = l^2c^2(-E')(Er)^{-1}$. This shows that the divergence is the dominant component of the motion where the oscillation frequency is sufficiently larger than ν_{cd}, and that the curl dominates where ν is small enough. Figure 1 displays ν_{cd} as a function of radial distance in the Sun. The full curve and the segment near the surface are for modes with $l = 1$; the two other short segments indicate the locations of parallel curves for the tenth and hundredth harmonics. We see that the curl dominates the motion of low-frequency modes (g-modes) in the solar core and that the divergence is dominant for all modes in the main solar convection zone. The curves cannot be drawn in a convection zone because λ_{cd} is a small, negative quantity there. The solar model used in this paper is No. 6 of Weymann and Sears (1965), joined smoothly in the deep photosphere to the Harvard-Smithsonian Reference Atmosphere (HSRA) (Gingerich *et al.* 1971).

Equation (16) is a rather general result, requiring no knowledge of how the oscillation is distributed with depth. The only properties of the mode which enter are its frequency and the principal index, l, of the spherical harmonic involved. According to (16), purely incompressible or purely irrotational motion is approached in the limits of very small or very large ν, respectively. These limiting forms help in visualizing the motion but give no immediate mathematical simplification because, setting either A_c or A_d to zero causes the other to vanish also (by eq. [16]), and no spheroidal motion exists. A single exception is known where oscillations completely free of divergence are possible; this is the textbook example of a homogeneous sphere (Robe 1965; Aizenman, Smeyers, and Weigert 1977).

III. RESULTS USEFUL IN ALL BUT THE LOWEST HARMONICS

a) p-Regions, g-Regions, and Mode Classification

For $l > 0$, it is very common to regard the gravitational perturbation W as negligible, whereupon the equations of motion are reduced to second order. The assumption appears to fail sufficiently close to the center of a star, where it can be shown that W makes a contribution to the momentum equation (2) comparable to the pressure gradient, the only other significant force in the region. However, the assumption is saved by the fact that the central stellar regions are of rapidly diminishing importance to an angular harmonic mode with $l > 1$, because the oscillation amplitude goes to zero at the origin. The traditional reasons for ignoring W in the outer regions of the star and for high harmonics were first stated by Cowling

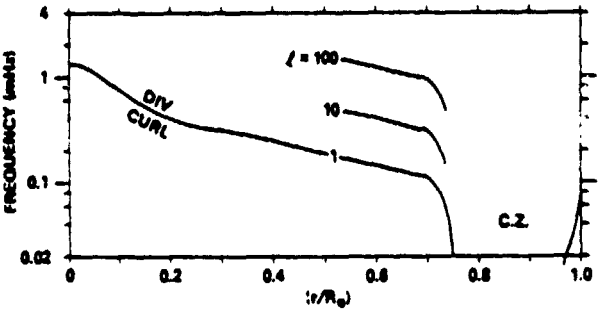


FIG. 1.—Curl-divergence frequency ν_{cd} separating zones in a star where the oscillatory motion is given predominantly by $\nabla \times v$ or mostly by $\nabla \cdot v$. A conventional solar model is used and the complete curve applies when the spherical harmonic index $l = 1$. Short segments are given of the parallel curves for $l = 10$ and 100. The divergence dominates for all frequencies of practical interest in the solar convection zone.

(1941). Numerical work—especially that of Robe (1968)—shows that the approximation, $W = 0$, causes little error except for the lowest radial harmonics and for the case $l = 1$. Excepting these few modes, the applicable equations are (8) and (9), with $w_1 = 0 = w_2$. They can easily be combined into a single second-order equation by eliminating either P or Q. I do this, keeping Q, and introduce the new radial variable,

$$u = \int dr\, D(\lambda - \lambda_b), \tag{17}$$

which is monotonic with r in any interval where $\lambda - \lambda_b$ does not change sign. In this variable, the equation reaches the irreducibly simple form of a wave equation,

$$\frac{d^2Q}{du^2} + \frac{Q}{S} = 0, \tag{18}$$

where $S = D^2c^2\lambda(\lambda - \lambda_b)(\lambda - \lambda_l)^{-1}$. A few attempts to sketch a function whose second derivative bears a constant sign relative to the function itself will prove that, if S is positive, the oscillation amplitude Q can have numerous nodes and extremes as u increases. If S is negative, Q must be a growing or decaying function of u with one node, at most. Equation (18) becomes the most compact statement of this sinusoidal or exponential behavior which Cowling discovered in the limit of very large or very small ν, and which heretofore has been proved by appeal to Sturm-Liouville theory. The radial momentum function P behaves similarly to Q because its wave equation also contains S as the sign-determining factor. The two frequencies controlling the sign of S are the buoyancy frequency ν_b and a horizontal acoustic frequency ν_l. The first is also called the Brunt-Väisälä frequency, and the second gives the time ν_l^{-1} in which sound travels one horizontal wavelength along a typical arc concentric with the star. Equations (4), (10), and (11) give the analytic forms. Scuflaire (1974), Unno (1975), and Osaki (1975) saw the utility of plotting these frequencies versus radial distance; Figure 2 is an analogous plot for the solar model. A solar

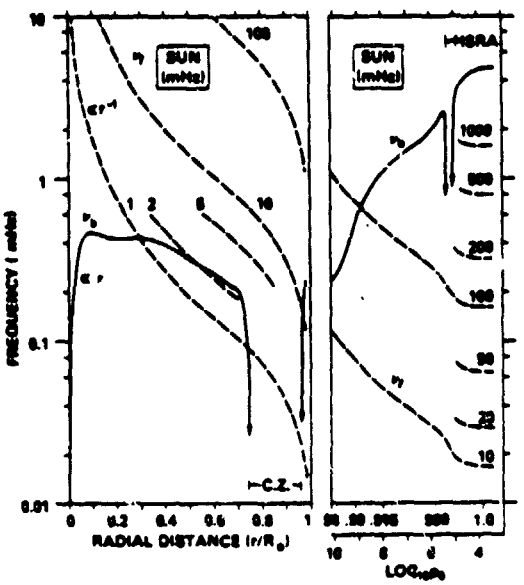


FIG. 2.—Two basic frequencies controlling the local characteristics of stellar pulsation modes. The solid curve gives the buoyancy frequency ν_b and a single dashed curve, labeled by the appropriate value of l, gives the horizontal acoustic frequency ν_l. Where the oscillation frequency lies well below both curves, the oscillation is like a classical g-mode. Where it lies well above both curves, the behavior is like a p-mode. The HSRA models the surface layers of the Sun and contains a density inversion causing the narrow trench in the ν_b curve near 0.999 $R_\odot$. The wider trench above 0.76 $R_\odot$ is caused by the solar convection zone.

convection zone above 0.7 $R_\odot$ and a density inversion in the visible layers near 0.999 $R_\odot$ cause λ_b to take on small, negative values in these layers. Near the origin, the figure shows functional dependences of r^{-1} and r which should hold in all stars with thermally supported, nonconvective cores. Since there is no oscillation frequency for which S is the same sign throughout the star, there is no mode everywhere possessing the features Cowling found in modes in polytropes in the limits, $\nu \to 0$ or, ∞.

Generalization of the Cowling classification to real stars is long overdue. The local properties of a mode depend crucially upon the size of two frequency ratios, $F_1 = (\nu/\nu_l)^2$ and $F_2 = (\nu/\nu_b)^2$, implicit in S. Only where both F_1 and $F_2 \gg 1$ does a stellar mode resemble Cowling's extreme p-mode, namely, primarily radial motion and relatively large fractional pressure fluctuations. Only where F_1 and $F_2 \ll 1$ is an oscillation like an extreme g-mode, with mainly horizontal motion and relatively small fractional pressure changes. Both extremes can occur in the same mode at different locations in a real star. Let us define a p-region (g-region) as a shell in which F_1 and F_2 are everywhere greater than (less than) 1. Both are regions where S is positive, and both are bounded by spheres where either F_1 or $F_2 = 1$. Scuflaire has shown that these regions are analogous to layers in a plane-parallel atmosphere where acoustic or gravity waves "propagate." Most of the kinetic energy of a typical linear mode is contained within a single S-positive region (§ IIIc, and Goosens and Smeyers 1974). This region may be thought of as a layer in which the oscillation is "trapped" by partially reflecting layers (S-negative) above and below. The boundary conditions can still be applied at the center and surface of the star, but numerical difficulties may arise when there is almost perfect reflection. Other locations have been used for the boundary (Dziembowski 1977; Osaki 1977) to reduce computer time and to treat cases with heavy damping outside the main S-positive region. A star has many more sets of pulsation modes than the conventional classification can accommodate if the star has many S-positive regions. This was discovered by Tassoul and Tassoul (1968) in an abstract star consisting of alternating convective and radiative layers. Since then, Goosens and Smeyers have studied the energy distribution among polytropic layers, and Shibahashi and Osaki (1976) have studied a massive star with molecular weight gradient in the core. They find a whole spectrum of modes associated with each S-positive region. Thus, a modern classification must state in which of these regions the energy is concentrated. Then the modes can be further labeled by the two integers, l and m, defining the angular harmonic, and by the radial harmonic number $|k|$, counting in consecutive integers from zero at the lowest radial harmonic for a given angular state. Modes with $k = 0$ are the f-modes of the Cowling nomenclature and are often exceptional in extending significantly over much of the star. They seem to have no true analog in the case $l = 0$ (see discussion of Fig. 1 in Cox 1976). Thus, it might be best to start with $k = 1$ when numbering the spherically symmetric cases. This would differ by one from the labels used by Cox and others, but would agree with the conventions of Christensen-Dalsgaard and Gough (1976) and Hill.

The four-dimensional classification (activated region, k, l, m) presented above can often be shortened. For example, m can often be dropped in slowly rotating stars because the quantity of interest depends only weakly on m. I use the additional convention of positive integers for k when a p-region has most of the energy, and negative integers when a g-region does. This is convenient in simple stars because the context often makes clear that there is only one p-region and one g-region of interest; then, reference to the region can be dropped also, leaving just the pair (k, l) for identification. A special difficulty in classification arises when the oscillation frequency of one mode, concentrated in a p-region, is very close to the frequency of another mode concentrated in a g-region. Aizenman, Smeyers, and Weigert (1977) showed that the two layers act like coupled oscillators, perturbing the frequency of both modes. Their graphs also show how a mode with a single name under the conventional nomenclature can gradually change its behavior from p-type to g-type to p-type, etc., as the star evolves. The new classification suggested herein gives the more appealing result of labeling a mode with certain physical properties by the same set of integers at most instants in the evolution of the star. Finally, our

scheme could also incorporate the "convective" solutions concentrated in negative-S regions for which the notation has been g_l^-. However, only under exquisitely controlled laboratory conditions do linear convective modes seem to be observed. In real stars, with the possible exception of semiconvection, self-excited convection seems to be a strongly nonlinear phenomenon to which our equations would rarely apply. I find the practice of classifying convection with the linear pulsation modes to be distracting.

To illustrate the use of Figure 2, consider a mode with energy concentrated in the main g-region of the Sun and $(k, l, m) = (-3, 2, m)$. Its oscillation frequency is about 0.24 mHz. Drawing a horizontal line at this frequency will show that the mode lies in a large g-region ($0.024 < r/R_\odot < 0.61$), a fairly big p-region ($0.64 < r/R_\odot < 0.98$), and negative-S regions elsewhere except in the thin shell near 0.999 containing the density inversion. Although this is conventionally called a g-mode (g_3^+, $l = 2$), Figure 2 shows that the extreme properties of a Cowling g-mode never occur because $F_2 \approx 1$ throughout the g-region. But p-type characteristics do occur because F_1 and $F_2 \gg 1$ in much of the main p-region. This overlap between the physical properties of what have been called p-modes and g-modes was discovered by Dziembowski (1971) in giant stars. It seems more useful just to number the modes with integers, k, as suggested above, and to start paying more attention to the local values of crucial ratios such as equations (16) and (24), giving the relative sizes of important physical quantities.

Figure 2 can also illustrate the origin of two types of modes, trapped in the outer layers of the Sun, which were originally suggested as explanations for the 5 minute oscillations ($\nu \approx 3$ mHz). Although both types were derived for plane-parallel layers extending to infinity, they locally approximate a global mode if their horizontal wavelength is short enough. The modes of Thomas, Clark, and Clark (1971) were trapped in the thin g-region above $0.999\ R_\odot$ which exists only if l is about 2000 or greater. These modes have horizontal wavelengths shorter than is now thought to be typical of the phenomenon. Ulrich (1970) studied other modes trapped in the p-region above $0.97\ R_\odot$ corresponding to $l \approx 400$. Ulrich's modes are a special case, in the limit of large l, of the global p-modes which Wolff (1972*b*) eventually suggested as the general explanation.

b) Formal Solution under the Slowly Varying Approximation

Except for low radial harmonics, the magnitude of a fluctuation usually changes much more rapidly with radial distance than does the equilibrium stellar model. In that case, the interior can be described as slowly varying, and its spatial derivatives higher than first order can be ignored. The wave equation (18) applies, and direct substitution will show that

$$Q = a_2 S^{1/4} e^{i\Phi} \tag{19}$$

solves (18) if the single term, $d^2/dr^2(S^{1/4})$, is ignored. The phase angle is defined by $\Phi = \int dr S^{-1/2}$ or, equivalently,

$$\Phi = \lambda^{-1/2} \int dr c^{-1} (\lambda - \lambda_l)^{1/2} (\lambda - \lambda_s)^{1/2} . \tag{20}$$

The constant a_2 cannot be determined by a linear theory. The orthogonal solution, derived in a similar way, is

$$P = a_1 S^{-1/4} e^{i\Phi} . \tag{21}$$

In other work, Smeyers (1968) derived asymptotic solutions in terms of Bessel functions and Dziembowski (1971, 1977) and Osaki (1977) derived sinusoidal forms in which equation (20) is approximated by $\lambda \to 0$ or $\lambda \to \infty$. The radial and angular components, P and Q, are typically 90° out of phase along the radius, as is known from exact integrations. We can see this roughly with the approximate solutions by substituting equations (19) and (21) into equations (8) and (9). Ignoring W and all derivatives of S, one gets $a_1 \approx i\lambda a_2$. Since a single integral (20) determines the solution $P(r)$, $Q(r)$, the fourth-order system of differential equations has been reduced to first order in this approximation.

There is a penalty for the simplification. Well known regular singularities in the complete system become manifest in solutions of the approximate problem. False cusps will appear as S changes sign. Fortunately, this affects only very thin shells because of the quarter power to which S is raised. Figure 3 shows the approximate solutions (*dashed curves*) and the exact solutions (*solid curves*) obtained much more laboriously from solution of the fourth-order system. The solar model was used. The solutions are for the modes $(k, l) = (-10, 2)$ and $(-3, 2)$, and each solution was normalized to unity at one of its extremes. The oscillation frequencies are also shown in Figure 3. The angular component Q contains the great majority of the kinetic energy in the tenth harmonic and about two thirds of it in the third harmonic. Thus, we have a good sample of the total motion by plotting only this component. The figure shows excellent agreement for $(-10, 2)$ throughout the g-region and even into the central core. The heights of extremes of the dashed curve agree to 1% with those of the exact solution except for the two closest to the center, where fluctuations in the gravitational potential are no longer negligible. Even for the third harmonic, where the slowly varying approximation must have considerably less validity, the dashed curve corresponds fairly well with the exact curve, as can be seen on the lower half of the figure. Toward the surface, the dashed curves are not shown because our approximations fail in the absence of nodes in these layers. Unlike these low-frequency modes, high-frequency modes in p-regions can have lots of nodes in the outer layers of a star. Then the slowly varying approximation can hold throughout the star, often remaining valid up to the photosphere.

Figure 3 shows a fairly constant phase difference of 12° between the upper curves and an average of about

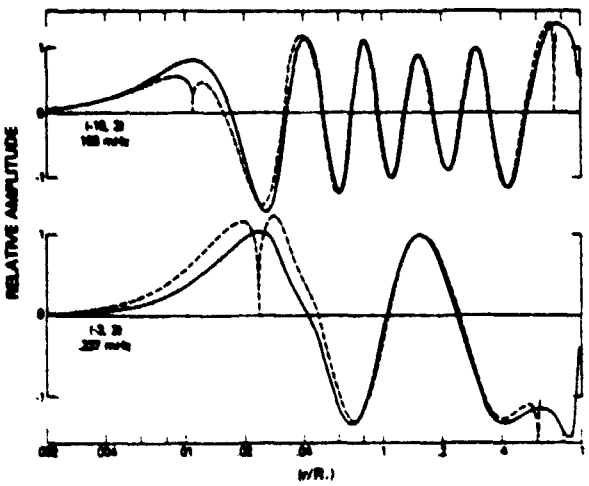


FIG. 3.—Oscillation amplitude Q, proportional to angular momentum per unit mass, is plotted against radial distance for two typical oscillation modes concentrated in the main g-region of the Sun. The dashed curves were obtained from simple, approximate formulae and agree very well with the exact solutions (*solid curves*) obtained by numerically integrating the 4th order differential equations of motion. At the boundaries of the g-region, the approximate solutions have unrealistic cusps which could be removed by a smooth joining of solutions across the singularity. The ordinate of each of the four curves has been normalized to unity at one of the extremes of the curve.

the same amount between the lower curves. Although one could artificially remove this from the approximate solutions with prior knowledge about the exact solution, the phase is not really a free parameter. The physics contained in equations (8) and (9) with $W = 0$ says that either P' or Q' should vanish where either ν_l or ν_0 equals the oscillation frequency. This means that the lower limit on the phase integral (20) must be chosen so that this condition is satisfied at one of these singular points. For the two modes on Figure 3, the outer boundary of the g-region was used. The inner boundary would not be as good a choice since W is not negligible there. To summarize, the simple solutions (19) and (21) apply to all but the lowest harmonics. They are valid at least in those parts of the star where the energy of motion is concentrated, and they always fail in thin shells near the singular points. (Bessel function solutions of Smeyers do not fail at these points.) Our solutions are expected to hold over almost all of the stellar volume for high frequency modes. Obviously, our simple solutions are realistic enough for many applications. They will be used below to derive an analytic expression for the energy distribution and asymptotic expressions for the oscillation frequencies.

c) *Energy Distribution*

The exact expression for K_s, the amplitude of oscillatory kinetic energy in a shell of unit thickness, is usually written as

$$\lambda r^2 \rho_0 [\xi_r^2 + l(l+1)\xi_a^2] .$$

In our momentum variables, it becomes

$$K_s = DP^2 + \lambda_l D^{-1} c^{-2} Q^2 . \tag{22}$$

The real parts of the approximate solutions (19) and (21) represent physical quantities and will be substituted into (22). But, before taking the real parts, a relation must be used to eliminate either a_1 or a_2, since there can be only one arbitrary scale factor. Using $a_1 \approx i\lambda a_2$, as given earlier, one gets

$$K_s = a_2^2(\lambda^3 D S^{-1/2} \sin^2 \Phi + \lambda_l D^{-1} c^{-2} S^{1/2} \cos^2 \Phi) . \tag{23}$$

Successive minima of K_s occur very close to locations where Φ has changed by an angle π because all other factors in (23) are slowly varying by assumption. The result of integrating (23) over such an interval is to replace $\sin^2$ and $\cos^2$ by their mean values, $\frac{1}{2}$, and to leave other factors unchanged. The ratio of the two terms in this integral is simply

$$\frac{\lambda}{\lambda_l}\left(\frac{\lambda - \lambda_l}{\lambda - \lambda_0}\right) = \frac{\text{Total K.E. in radial motion in the shell}}{\text{Total K.E. in angular motion in the shell}} . \tag{24}$$

The fraction on the right reflects the fact that the two terms in (23) are proportional to, respectively, radial and angular components of motion. Any shell referred to in (24) is bounded by successive locations where K_s is a minimum. A very large oscillation frequency causes the motion to be primarily radial, and a very small one causes horizontal motion as Cowling found in polytropes. But equation (24) is much more general and precise, holding for any value of ν, for any stellar interior model, and for any portion of a star where the static interior varies slowly with radius compared to the magnitude of a fluctuation.

It also follows from (23) that the total kinetic energy in neighboring shells must be almost equal. But the equality holds far beyond nearest neighbors and is important enough to demonstrate with a numerical solution of the full fourth-order system (eqs. [1]–[3] and Poisson's equation). The curve on Figure 4 shows K_s as a function of radial distance for the solution $(-10, 2)$ discussed earlier. Use of the oscillation frequency of 0.108 mHz and Figure 2 define the location of the p- and g-regions shown on Figure 4. We see that kinetic energy lies almost entirely in the large solar g-region. The area under the curve between successive minima is the total kinetic energy in an antinodal shell. The area must be found using a linear distance scale, of course, and the results are given on the histogram. It shows that energy is distributed rather equally among the antinodes, except for a boundary effect at each end. The five centermost antinodes have total energies differing by less than 1% from their mean! Even when those at the ends are included, equipartition of energy holds to $\pm 20\%$, even though the volume and density of the shells

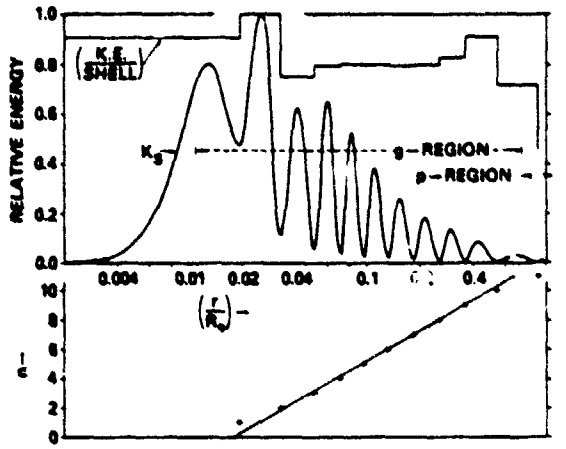


FIG. 4.—Kinetic energy per shell $K_s(=4\pi r^2 \cdot \frac{1}{2}\rho v^2)$ as a function of relative radial distance for the 10th harmonic referred to on Fig. 3. At least 95% of the energy lies in the main solar g-region. The total energy associated with each antinode (each maximum of K_s) is shown on the histogram. Equipartition of energy holds to $\pm 1\%$ for the five antinodes in the middle and holds with increasing scatter as one moves toward each edge of the g-region. Each minimum of K_s is spaced almost equally from its neighbors on the logarithmic distance scale. The lower panel tests this more critically by plotting the location of each minimum against n, the number of the minimum, counting from zero at the origin.

differ greatly. Equipartition also holds in high-frequency modes in the main p-region of the Sun, as could be proved roughly by measuring the small graph given by Wolff (1972*b*) for the mode (10, 10) for which $\nu = 2.05$ mHz. Finally, an unknown amount of energy lies near the surface on Figure 4, and the curve does not extend the last few hundredths of the radius. It is not realistic to calculate this with adiabatic equations, such as ours, which ignore convective and radiative losses. For example, strong coupling might occur with convection above $0.98\ R_\odot$, where many solar models show active convection with cell-overturning times comparable to the 3 hour period of this pulsation mode. Equipartition of energy among antinodes should hold very well for all high harmonic modes of stellar pulsation and should hold roughly even for low harmonics. It has escaped notice until now because it was concealed by boundary effects in the low harmonics which it has been customary to study.

d) Distribution of Nodes and Eigenvalues

Each minimum of the kinetic energy function on Figure 4 appears to be equidistant from its neighbors on the logarithmic distance scale. This is shown quantitatively on the bottom half of the figure where the location, r_n, of the nth minimum is plotted against n. Indeed, all the points except those at the ends lie near the straight line given by

$$r_n = be^{an}, \qquad (25)$$

where the two constants are chosen for best fit. A good fit, applying to all the solar cases seen so far ($l = 2, 5, 8; k = -5, -10$), is obtained with the choice $a \approx (\frac{1}{2} - k)^{-1} \ln (r_0/r_i)$ where r_0 and r_i are the outer and inner boundaries of the g-region. The constant, $\frac{1}{2}$, adjusts for boundary effects. It was empirically determined and may require an addition of $\pm\frac{1}{2}$ when the g-region is bounded on one side by $\nu = \nu_i$. The scale factor b is less regular but is always roughly equal to r_i. Using equation (25) one can write down the approximate location of each kinetic energy minimum without performing any numerical integrations. Each such minimum is practically coincident with a zero of Q or ξ_r for most low-frequency modes. This exponential distribution of nodes has a range of validity at least as broad as the range over which the following proof holds. Start with the solutions of § III*b* and further require that $\lambda \ll \lambda_i$. Then the phase integral (20) simplifies to $-l^* \lambda^{-1/2} \int dr r^{-1} (\lambda_b - \lambda)^{1/2}$. For brevity, the last factor will be treated as constant in the range $0.04 < r/R_\odot < 0.5$, although a proof could still be carried out analytically under the next higher approximation, $(\lambda_b - \lambda)^{1/2} \propto r^{-\epsilon}$ and $\epsilon = 0.1 \ll 1$, in the Sun. An integration between any two minima, r_n and r_{n+q}, lying in this range, must give a phase change of $\Delta\Phi = -\pi q$, so that equation (20) reduces to $(r_{n+q}/r_n) = e^{\alpha q}$, where $\alpha = \pi \lambda^{1/2} [l(l+1)\langle \lambda_b - \lambda \rangle]^{-1/2}$, a constant. Since this shows that the minima should be distributed exponentially along the radius, it proves the empirical rule (25).

An analogous rule exists for the nodal distribution of high-frequency modes in p-regions. As others have already shown, it depends on the sound travel time,

$$\tau_n = \int_{r_n}^{r_{n+1}} dr c^{-1}, \qquad (26)$$

between two neighboring nodes. Wherever $\lambda \gg \lambda_i$ and λ_b, the phase integral reduces to $\tau_n = (2\nu)^{-1}$. This says that the nodes in p-regions tend to place themselves along the radius so that a sound wave can travel between them in one half an oscillation period—a result familiar in the laboratory for standing waves in an unstratified medium.

Surprisingly acceptable values for the oscillation frequencies can also be obtained under the slowly varying approximation. First, consider several limiting forms. When the frequency is very low, the innermost portions of a g-region contribute only a few percent to a phase integral carried out over the full range of the g-region. Then, the exact integral is adequately estimated by choosing some effective lower limit r_e which is about $0.04\ R_\odot$ in the Sun, and by adopting a mean for ν_b. The mean should typify the buoyancy frequency where most of the antinodes are located, and 0.4 mHz is not a bad choice for the Sun (cf. fig. 2). Equation (20) then becomes $\Delta\Phi \approx -l^* \bar{\nu}_b \nu^{-1} \ln (r_0/r_e)$, and, since $\Delta\Phi \approx k\pi$ as we shall see again below, the limiting form for small oscillation frequencies is

$$\nu = \frac{l^* \bar{\nu}_b}{-k\pi} \ln \left(\frac{r_0}{r_e}\right) \quad \text{(for small } \nu\text{)}.$$

Large frequencies typify modes concentrated in p-regions. To estimate the frequencies, carry out the

phase integral over the full extent of the p-region, obtaining $\Delta\Phi = \lambda^{1/2}\tau$, where τ now represents the sound travel time across the p-region. The radial order number, k, should be related to the total phase shift again by $\Delta\Phi \approx k\pi$. With this, one obtains

$$\nu \approx \frac{k}{2\tau} \quad \text{(for large } \nu\text{)}.$$

More elaborate formulae have been derived (for example, Vandakurov 1968) by forcing the interior model to obey a power law near the surface and the center. Our formulae agree rather well with a large array of exact solar frequencies plotted as Figure 2.2 by Hill (1977). Across one third of his plot (where $k > l$) ν is almost independent of l. This agrees with our formula for large ν because τ approaches a constant for large ν as the lower boundary of the p-region enters the deeper layers where sound travel is so rapid. The formula for small ν given above applies at solar frequencies less than 0.1 mHz. Although this is just off scale on the Hill plot, behavior at the bottom of the plot is approaching a simple proportionality, $\nu \propto l^{*}(-k)^{-1}$, consistent with the above formula.

There is a third realm where the oscillation frequencies approach a constant limit independent of all three harmonic indexes, k, l, m. The limit is the highest value of ν_b in the g-region and it applies only to g-modes:

$$\nu \to \text{MAX}\,(\nu_b) \quad \text{(for modes in } g\text{-region with } l \gg |k|)\,.$$

Shibahashi and Osaki (1976) describe it as the saturation effect and note that it occurs when the horizontal wavelength becomes sufficiently short compared to the radial wavelength. It is best displayed on the Hill plot and could also be extracted from (20) with the extra condition that $\partial\nu/\partial l \geq 0$.

When ν falls into none of the above categories, it is necessary to numerically integrate (20) for a series of trial frequencies. The integral can be carried out over the full radius of the star, but only the real parts, $\Delta\Phi$, extending over the complete range of a g-region or p-region are used below. For the solar case $l = 2$, values of the integral for about a dozen frequencies were plotted on Figure 5 and then replaced by smooth curves representing all the results. Eigenfrequencies of the problem should be separated by nearly integer increments of $\Delta\Phi/\pi$ on the plotted curves. The plus signs are exact oscillation frequencies of the complete fourth-order system plotted against radial harmonic index, k. It is clear that the curves asymptotically approach the distribution of exact frequencies. Equally good results have been seen for the cases $l = 5$ and $l = 10$. In all these cases, the frequency of the fundamental mode ($k = 0$) lies close to the point on the central axis equidistant from the curves for the p- and g-regions. An interpolation scheme for the other low harmonics on the plot may not be too difficult to derive if needed.

A set of curves as on Figure 5 represent an economical way to survey the pulsation frequencies of a new stellar model with the added advantage that the condition, $k \approx \Delta\Phi/\pi$, removes ambiguity over the identification of the radial harmonic number. Ambiguity has been an occasional problem for many years in the more centrally condensed stars. Up to now, it could only be removed by computing a long series of exact solutions for consecutive harmonics (Robe 1968; Ledoux 1974).

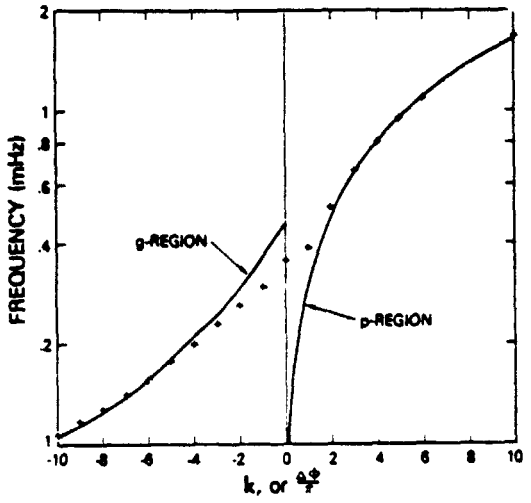


FIG. 5.—Exact solar oscillation frequencies for the case $l = 2$ are plotted (+) against radial harmonic number, k. Except for the low harmonics, the frequencies follow the solid curves which represent the elementary phase integral (eq. [20]) divided by π. When the curves have an integer value on the abscissa, their ordinate gives the approximate value of the true oscillation frequency. The point on the central axis equidistant from the two curves lies close to the exact frequency of the fundamental mode ($k = 0$) in this case and in the two other cases tested ($l = 5$ and $l = 10$).

IV. CENTRAL AND POLAR SINGULARITIES

A standing oscillation can be regarded as dividing a star into cells, each containing one antinode. Those cells near the center of the star and near its spin axis will be shown to have more oscillatory kinetic energy per unit volume,

$$K_v = r^{-2}(DP^2|Y|^2 + D^{-1}Q^2|\nabla Y|^2)\,, \qquad (27)$$

than the other cells. When r^2K_v is integrated over a thin spherical shell of unit thickness, it reduces to K_s (eq. [22]). The cell boundaries are taken as the surfaces where r^2K_v is a minimum. The simple nature of the function K_v gives one maximum, or antinode, in each cell. We will consider only cases where one term in (27) is negligible, but this will include almost all modes. That is true because the ratio of equation (24) is usually $\gg 1$ or $\ll 1$ over most of the region in which the oscillation is concentrated, except for the lowest radial harmonics. The cell boundaries reduce to the surfaces where $|PY|^2 = 0$ for the high frequency branch (p-modes) and to minima of $|Q\nabla Y|^2$ for the low frequency branch (g-modes). In each type of mode, no fluid is carried across the cell boundaries by the dominant velocity compone f. eq. [7]).

Only the small orthogonal component can move fluid out of the cell. Thus, rigid walls could suddenly be placed around a typical cell, and the fluid would continue to oscillate substantially in the same manner for a few cycles. In other words, the standing wave has divided the motion into an array of relatively self-contained cells, each of which has about the same natural frequency.

The number of these cells is rather large. Even a modest harmonic like $|k| = l = 5$ divides the star into ≈ 76 cells, as is now shown. For the high-frequency cases, zeros of Y_l^m are of interest, and they divide a spherical surface into exactly $(l - |m| + 1)(2 + \delta_{m0})$ areas. When $m = 0$, $\delta_{m0} = 1$ and in other cases, $\delta_{m0} = 0$. For compactness, the number of areas can be averaged over the states $m = 0, 1, 2, \ldots l$, giving $\frac{1}{2}[(l + 1)^2 + 2] \equiv N_l$ areas on the sphere for a typical standing mode of principal index, l. The low-frequency cases are somewhat less regular, but computations show that minima of $|\nabla Y|^2$ also divide the sphere into N_l areas with minor deviations $\sim l^{-1}$. There are $(|k| + 1)$ significant antinodes along a radius, so that the total number of cells in the star is $\approx (|k| + 1)N_l$, which rises rapidly with increasing harmonic number toward the asymptote, $\frac{1}{2}|k|l^2$.

a) Enhancement Near the Center

The cell volume in low-frequency modes is very much smaller near the stellar center than in its outer regions. For example, the outer cells of the mode on Figure 4 have $\sim 10^5$ times the volume of those near the center. Recalling the rough equipartition of energy among cells, it is obvious that kinetic energy density is enormously greater in centrally located cells of g-modes for which $-k \gg 1$. In fact, K_v might well be significant to the central structure and evolution of stars whose surfaces are barely pulsing. This possibility will be applied to the "low" value of the solar neutrino flux in a paper under preparation. In the shell of width Δr containing one layer of antinodal cells, the kinetic energy is $\sim \frac{1}{2}\lambda r^2 \rho_0 |\xi|^2 \Delta r$, evaluated at an antinode. This comes from the expression preceding equation (22) and is roughly constant for all shells. Therefore, we can equate its values for the inner (i) and outer (o) shells of the g-region and write the result in the form,

$$J_i^2 \sim J_o^2 (\rho_0 r^4 \Delta r)_o / (\rho_0 r^4 \Delta r)_i \,. \qquad (28)$$

The quantity J is the displacement $|\xi|$ at an antinode, divided by the typical horizontal dimension of the cell, $(4\pi r^2 N_l^{-1})^{1/2}$. Also, the ratio of the factors $r^4 \Delta r$ can be taken as $r_o^5 r_i^{-5}$ by use of equation (25) in the limit of large $|k|$. To estimate the size of (28) it is adequate to take for r_i and r_o the inner and outer boundaries of the g-region. For the Sun, $r_i \approx R_\odot l^* \times (-20k)^{-1}$ and $r_o \approx 0.7\, R_\odot$ by use of Figure 2 and the asymptotic expression for small eigenvalues. At these locations, $\rho_0 \approx 160$ and $0.1\ \mathrm{g\ cm^{-3}}$, so that (28) becomes $J_i^2 \approx J_o^2 (-3.2k/l^*)^5$. The fact that k is raised to the fifth power shows, again, why g-modes of high radial harmonic must have relatively large amplitudes in the center of a star.

There is no mathematical limit to how large $-k$ can become, but turbulence and radiation provide physical limits. Severe damping from these causes should set in first at the innermost antinode. Let us assume that turbulence is already large when the oscillatory displacement approaches the scale of the flow pattern (that is, $J_i \to 1$). Further assume that J_o is only $\sim 10^{-7}$, which (for small l) is ~ 100 times smaller than fluctuations seen at the solar limb by Hill's group. Then by (28), as evaluated above for the Sun, the turbulence limit has already been reached when $-k > 200/l^*$. Dziembowski (1971) numerically integrated the fourth-order equations and found heavy radiation damping in the core for radial harmonic numbers of a few thousand. This was for giant stars, but an order of magnitude estimate for the solar interior gives the following results: Treating the center of the Sun as a homogeneous, ideal gas undergoing compression of high radial wavenumber, k_r, such that $k_r r \gg 1$, I computed t_d, the radiative decay time (Cox 1974, § 5.5), under optically thick conditions. The numerical result can be written in the convenient form,

$$2\pi\nu t_d \approx \left(\frac{190}{-k}\right)^5 \left[\frac{l(l+1)}{6}\right]^{1/2} \left(\frac{T}{T_0}\right)^{3/2} , \qquad (29)$$

where T is the perturbation in the mean temperature T_0. Although t_d depends only on the (high) radial harmonic number in this estimate, l also enters (29) by its influence on ν. Radiation damping is severe if $2\pi\nu t_d \leq 1$, which (by eq. [29]) will occur before $-k$ is as large as 190 for $l = 2$ modes. For $l \geq 2$, the radiation damping limit is reached before the turbulence limit so that, in summary, no one should expect the analytic formula (28) for central enhancement to be valid in solar g-modes when the radial harmonic number exceeds about $150\,(l^*)^{0.2}$. Such high harmonics will require a complete nonlinear treatment.

b) Enhancement Near the Rotation Axis

The angular distribution of kinetic energy among cells located at the same distance from the stellar center is now computed. This will be a purely geometric effect, independent of the details of the interior model, provided that the star performs slow, uniform rotation in the layers where the oscillation is concentrated. This condition guarantees that the centrifugal forces are small compared to gravity and Coriolis forces, and that the star is essentially spherical. The total kinetic energy in a cell is the volume integral of K_v. The latter contains a radial integral which is identical for any of the cells under discussion. Canceling this out, one finds that the relative energy in any cell at the same distance from the stellar center is proportional to a surface integral over the solid angle subtended at the center of the star by the cell. The integrand is $|Y|^2$ for the high-frequency modes and $|\nabla Y|^2$ for the low-frequency modes. These were carried out numerically over θ and analytically over

TABLE 1

RATIO OF KINETIC ENERGY DENSITY* IN HIGH-FREQUENCY MODES

	m											
l	0	1	2	3	4	5	6	7	8	9	10	AVERAGE OVER m
10	8.91	5.31	3.80	4.43	2.30	1.83	1.53	1.21	1.03	1.00	1.00	2.94
9	7.95	4.87	3.39	2.61	2.02	1.66	1.30	1.09	1.00	1.00	...	2.69
8	7.22	4.28	3.08	2.28	1.82	1.39	1.15	1.00	1.00	...	...	2.58
7	6.23	3.85	2.65	2.04	1.51	1.23	1.00	1.00	...	...	...	2.44
6	5.53	3.25	2.35	1.67	1.33	1.00	1.00	...	...	...	...	2.30
5	4.50	2.84	1.88	1.46	1.00	1.00	...	...	...	...	...	2.11
4	3.85	2.19	1.63	1.00	1.00	...	...	...	...	...	...	1.93
3	2.73	1.85	1.00	1.00	...	...	...	...	...	...	...	1.65
2	2.18	1.00	1.00	...	...	...	...	...	...	...	...	1.39
1	1.00	1.00	...	...	...	...	...	...	...	...	...	1.00
0	1.00	...	...	...	...	...	...	...	...	...	...	1.00

* Polar cell/equatorial cell.

ϕ for all 66 spherical harmonic functions with $l \leq 10$ and non-negative m. This showed that the cells of a given spherical harmonic mode had about the same kinetic energy except for a unique group close to the polar axis which could be either higher or lower than the rest. In contrast to this, kinetic energy per unit volume was found to be a monotonically increasing function of latitude, always largest in a "polar cell" (one touching or including the pole).

Tables 1 and 2 summarize the main findings for the high- and low-frequency modes. Entries in the main part of the table are the ratio of kinetic energy density in any one of the polar cells to that in an "equatorial cell" (one touching or including the equator). These ratios range from 1 to 8.91 for the modes studied, and tend to larger values at the higher latitudinal harmonics. Ratios of exactly 1 mark low harmonics where each polar cell is coincident with an equatorial cell. The concentration of energy density toward the poles is less for the low-frequency modes and would be reduced somewhat further if the following consideration were included. For most of the cases $1 \leq m \leq 4$, one more minimum appears close to each pole in $|\nabla Y|^2$ than in $|Y|^2$. The extra minimum is usually weak and of questionable physical significance. If it were to be ignored in defining the low-latitude boundary of the polar cell, the corresponding entries in Table 2 would be reduced; for example, the largest entry would go from 2.58 to 2.15.

For a physical system which excites standing waves of all m-values to roughly equal amplitudes, row averages of the tables can be of interest. These appear at the right of each table and show that typical polar cells can have 2 or 3 times the energy density of low-latitude cells. Such a polar enhancement may drive large-scale convection cells at each pole of some pulsing stars. This was mentioned by Wolff (1974*b*) who first found the polar enhancement in a less direct, but mathematically equivalent, manner. He suggested applying the large scale convection to the Sun's differential rotation and the qualitative differences long noted between its polar and lower-latitude phenomena.

V. MODE COUPLING

Up to this point, no mode considered has been coupled to any other mode because the equations of motion used have all been linear. But there are always nonlinear effects operating in real stars which can

TABLE 2

RATIO OF KINETIC ENERGY DENSITY* IN LOW-FREQUENCY MODES

	m											
l	0	1	2	3	4	5	6	7	8	9	10	AVERAGE OVER m
10	5.37	6.69	3.98	2.89	2.27	1.65	1.35	1.16	1.02	1.00	1.00	2.58
9	4.91	5.68	3.66	2.56	2.03	1.43	1.22	1.08	1.00	1.00	...	2.46
8	4.29	5.19	3.17	2.31	1.72	1.29	1.00	1.00	1.00	...	...	2.33
7	3.89	4.61	2.81	1.94	1.53	1.00	1.00	1.00	...	...	...	2.22
6	3.26	4.08	2.32	1.70	1.00	1.00	1.00	...	...	...	...	2.05
5	2.38	3.22	2.11	1.25	1.00	1.00	...	...	...	...	...	1.83
4	2.20	2.70	1.36	1.03	1.00	...	...	...	...	...	...	1.66
3	1.83	1.66	1.06	1.00	...	...	...	...	...	...	...	1.39
2	1.00	1.21	1.00	...	...	...	...	...	...	...	...	1.07
1	1.00	1.00	...	...	...	...	...	...	...	...	...	1.00
0	1.00	...	...	...	...	...	...	...	...	...	...	1.00

* Polar cell/equatorial cell.

couple the modes. We now consider some possible effects of weak coupling. The interactions are assumed to be weak enough that the spatial and temporal behavior of each linear mode is only slightly altered. However, our attention will shift from the linear mode to the properties of a sum of coupled modes. For decades, it has been of interest to inquire whether there might be two low harmonic modes in a given star with almost the same oscillation period. This could give a beat period that might be observable. Such an elementary possibility is only one of many. Three promising ways to couple are mentioned below, and each has the effect of simplifying the expected observational data.

In the stellar core, there are layers in which nuclear processes involving a particular element are especially effective in driving pulsation modes. For example, Unno states that the $^3He + {}^3He$ reaction in the Sun contributes most toward driving solar g-mode oscillations. Modes which have antinodes in layers where this reaction is strongest may well be excited to larger amplitudes than other modes. Furthermore, since this reaction is a highly nonlinear function of temperature, there will be a strong advantage in total energy released if those modes with antinodes in this layer and having similar oscillation periods can lock their periods together so that they are identical. Then a stronger rms amplitude is produced than if each mode pulses at an independent rate. It is quite possible that this mechanism operates in some stellar cores, reducing the vast array of linear oscillation periods to a small number of nonlinearly coupled periods.

Another possibility is based on the fact that all modes in g-regions with energy primarily in horizontal motion will rotate at nearly the same rate if they have the same harmonic index, l. Wolff (1974*a*) pointed this out and suggested that nonlinear coupling could lock similarly rotating modes together in the configuration having the largest possible rms amplitude. This would be their most stable configuration under appropriate nonlinear driving. Each group of locked modes must then rotate like a rigid body at a rate depending only on l and the rotation rate of the stellar mass. The precise rotation rates expected for the coupled modes have since been detected (Wolff 1976) in two independent kinds of solar data.

The third possibility operates near the highest frequency permitted for modes in the main g-region—that is, for frequencies near the broad maximum of ν_0 on Figure 2. Near this frequency, there is a great crowding together of all modes for which $l \gg |k|$. This was mentioned in § III*d*. It is quite unlikely that modes so densely situated along the frequency axis can maintain their independence in any real star which excites many such modes. If the modes do not couple spatially, in a way analogous to that in the previous paragraph, then they will almost certainly couple according to oscillation frequency. Observers of multiperiodic stars should be alert for above-average oscillatory power at a frequency in this vicinity due to the combined motion of many g-modes with high values of $|l/k|$.

All three mechanisms just described operate on the same principle: a nonlinear driving mechanism favors the concentration of oscillatory power into small volumes at large amplitude rather than into larger volumes at necessarily smaller mean-squared amplitude. This was stated intuitively and then supported by a numerical example by Wolff (1974*a*). It is correct as far as it goes but must eventually be combined with a statement about damping mechanisms. Presumably, the effective nonlinear power governing the damping must be smaller than that governing the driving in at least one portion of the star, or the whole effect will go in the opposite direction toward evenly distributed amplitudes.

VI. SUMMARY

Oscillatory kinetic energy is distributed almost equally among the antinodes of a linear stellar pulsation mode, with modest deviations for antinodes lying near physical or geometric boundaries. But the various antinodes can occupy greatly different volumes. This causes the kinetic energy density to be very much larger than average at antinodes near the stellar center and several times larger than average at antinodes near the spin axis of a slowly rotating star. These two unique zones could reasonably be suspected of also being zones of enhanced thermal dissipation. The dissipation may occasionally be strong enough to drive convection where it is not otherwise expected.

Without ever performing a numerical integration, values of many of the oscillatory quantities can be estimated, provided that the gravitational perturbation is negligible—that is, provided that $l > 1$ and that $|k|$ is not too close to 1. The oscillation frequencies are simple functions of k and l in the limits of high and low frequency, and they approach a constant value for all modes concentrated in g-regions when $l \gg |k|$. Given a rough value for the frequency, one can know the ratio of kinetic energy in radial and angular components of the motion (eq. [24]), the location of the nodes (eq. [25] and following), and the relative amplitude at almost all antinodes. The last condition follows from the fact that the same total kinetic energy is associated with almost every antinode while the volume it occupies is known from the location of the nodes.

If a certain phase integral (eq. [20]) is numerically evaluated at some representative frequencies in the range of interest, almost all the oscillation frequencies of the star can be read from simple graphs like Figure 5. The phase integral also permits one to write down elementary expressions for the eigenfunctions, and they agree very well with exact solutions (Fig. 3). Since the elementary expressions depend on a slowly varying approximation for the stellar interior, they fail in the outer envelope for low-frequency modes but would seem to hold up to the photosphere for most high-frequency modes.

For all modes and all locations, a simple, precise condition was found giving the fraction of the velocity field in divergence and the fraction in curl (eq. [16] and

Fig. 1). In the limit $\nu \to \infty$, the separate conditions (16 and 24) on the velocity and energy distributions imply a motion first found by Cowling in which pressure fluctuations and radial motion are dominant. The exact fourth-order equations governing linear adiabatic modes were presented in perhaps their most compact form using variables similar to those used by Ledoux and Walraven for the approximate second-order system.

Even though an oscillation mode occupies the entire star and must satisfy physical boundary conditions at the center and surface, its essential character and oscillation period is usually determined in a single zone of the star where its energy is concentrated. Scuflaire first defined the physical significance of these zones (called *p*-regions and *g*-regions herein). A classification of the modes based on these regions seems required. Finally, three methods by which the large number of linear modes might plausibly couple to produce simpler arrays were suggested in § V.

I thank an anonymous referee and P. Smeyers for many constructive comments on the first version of this paper.

REFERENCES

Aizenman, M. L., and Smeyers, P. 1977, *Ap. Space Sci.*, **48**, 123.
Aizenman, M. L., Smeyers, P., and Weigert, A. 1977, *Astr. Ap.*, **58**, 41.
Ando, H., and Osaki, Y. 1975, *Pub. Astr. Soc. Japan*, **27**, 581.
Christensen-Dalsgaard, J., and Gough, D. O. 1976, *Nature*, **259**, 89.
Cowling, T. G. 1941, *M.N.R.A.S.*, **101**, 367.
Cox, J. P. 1974, *Rept. Progr. Phys.*, **37**, 563.
———. 1976, *Ann. Rev. Astr. Ap.*, **14**, 247.
Deubner, F.-L. 1975, *Astr. Ap.*, **44**, 371.
Dziembowski, W. 1971, *Acta Astr.*, **21**, 289.
———. 1977, *Acta Astr.*, **27**, 95.
Fitch, W. S. 1973, *Ap. J. (Letters)*, **181**, L95.
Gingerich, O., Noyes, R. W., Kalkofen, W., and Cuny, Y. 1971, *Solar Phys.*, **18**, 347.
Goosens, M., and Smeyers, P. 1974, *Ap. Space Sci.*, **26**, 137.
Hill, H. A. 1977, *The New Solar Physics*, ed. J. A. Eddy (Washington: AAAS), Chap. 1.
Hill, H. A., and Stebbins, R. T. 1975, *Ann. N Y Acad. Sci.*, **262**, 472.
Hill, H. A., Stebbins, R. T., and Brown, T. M. 1976, in *Atomic Masses and Fundamental Constants*, Vol. 5, ed. J. H. Sanders and A. H. Wapstra (New York: Plenum).
Ledoux, P. 1974, in *IAU Symposium No. 59, Stellar Instability and Evolution*, ed. P. Ledoux *et al.* (Dordrecht: Reidel).
Ledoux, P., and Walraven, Th. 1958, in *Handbuch der Physik*, **51**, 353.
Osaki, Y. 1975, *Pub. Astr. Soc. Japan*, **27**, 237.
———, Y. 1977, *Pub. Astr. Soc. Japan*, **29**, 235.
Rhodes, E. J., Ulrich, R. K., and Simon, G. W. 1977, *Ap. J.*, **218**, 901.
Robe, H. 1965, *Bull. Acad. Roy. Belg. Cl. Sci.*, **51**, 598.
———. 1968, *Ann. Ap.*, **31**, 475.
Scuflaire, R. 1974, *Astr. Ap.*, **36**, 107.
Shibahashi, H., and Osaki, Y. 1976, *Pub. Astr. Soc. Japan*, **28**, 199.
Smeyers, P. 1968, *Ann. Ap.*, **31**, 159.
Tassoul, M., and Tassoul, J. L. 1968, *Ann. Ap.*, **31**, 251.
Thomas, J., Clark, P., and Clark, A. 1971, *Solar Phys.*, **16**, 51.
Ulrich, R. K. 1970, *Ap. J.*, **162**, 993.
Unno, W. 1975, *Pub. Astr. Soc. Japan*, **27**, 81.
Vandakurov, Yu. V. 1968, *Soviet Astr.—AJ*, **11**, 630.
Weymann, R., and Sears, R. L 1965, *Ap. J.*, **142**, 174.
Wolff, C. L. 1972*a*, *Ap. J.*, **176**, 833.
———. 1972*b*, *Ap. J. (Letters)*, **177**, L87.
———. 1974*a*, *Ap. J.*, **193**, 721.
———. 1974*b*, *Ap. J.*, **194**, 489.
———. 1976, *Ap. J.*, **205**, 612.

CHARLES L. WOLFF: Code 912, Goddard Space Flight Center, Greenbelt, MD 20771

NASA Conference Publication 2076 – Fourth NASA Weather and Climate Program Science Review, January 24-25, 1979, NASA/GSFC, Greenbelt, MD.

Paper No. 65

LONG-TERM MIGRATION OF THE SOLAR SECTOR STRUCTURE

C. L. Wolff and D. F. Heath, *Goddard Space Flight Center, Greenbelt, MD*

INTRODUCTION

Periodicities are often seen in weather records and immediately dismissed as not due to solar variability because the period is not a simple multiple of 27 days or 11 years. This is wrong because solar observations of the last few centuries show numerous other periodicities ranging from many months to many years. A theory is now under development which can predict the values of these periods and which will put this subject on a sound physical basis if the theory continues to be successful. Weather records can then be analyzed in a more comprehensive and respectable manner because the variable solar input will be more rigorously modelled.

The magnetic sector boundaries on the sun and in the solar wind have a high correlation with winter low pressure systems on earth. Roberts and Olson (1973) and Wilcox, et al. (1974) found that the vorticity-area index typically declines by about 10% during several days centered on the time when a sector boundary sweeps past the earth. The physical connection between these two events is not established. A whole group of other correlations exists between weather and solar activity (Goldberg and Herman, 1978). Again, much of the detailed physics is in doubt. Progress in understanding these phenomena is slow, partly because there are a very large number of possible explanations--each with its free parameters--and a relatively modest number of truly independent observational results. This paper is part of a continuing attempt to reduce the degrees of freedom available to sun-weather theories and, perhaps, to discover some more fundamental inputs to those theories. In what follows, we present new evidence that BOTH the sector structure and solar activity levels can be understood as being under the influence of the same regular, internal solar mechanism. Because of this, long-term predictions of each phenomenon may ultimately be possible.

IV. TROPOSPHERE BRANCH

Paper 36

Reprinted from Preprint Volume: Eleventh Conference on Severe Local Storms, Kansas City, Missouri, October 2-5, 1979. Published by the American Meteorological Society, Boston, Mass.

DETECTION OF SEVERE THUNDERSTORMS USING SHORT INTERVAL GEOSYNCHRONOUS SATELLITE DATA

Robert F. Adler

Laboratory for Atmospheric Sciences
Goddard Space Flight Center
National Aeronautics and Space Administration
Greenbelt, Md.

Douglas D. Fenn

GE/MATSCO
Beltsville, Md.

1. INTRODUCTION

The purpose of this paper is to discuss the potential and limitations of using short interval (3-7 minutes) geosynchronous, infrared (IR) data to ascertain thunderstorm intensity, and therefore indirectly detect severe thunderstorms. Geosynchronous satellites such as those from the SMS/GOES series view the area of concern (e.g. the midwest of the United States) with one sensor and with essentially a constant viewing angle. With this factor and the better time resolution provided by operating the satellite in the rapid-scan or short-interval mode one has the tools to attempt to estimate relative thunderstorm intensity. Assuming that the intensity of convection is correlated with the occurrence of severe weather one can use the satellite-based parameters to detect severe thunderstorms.

In this study thunderstorms on four case study days (May 6, 1975; April 24, 1975; March 20, 1976; May 20, 1977) are analyzed over selected areas and time periods using SMS/GOES data (see Table 1), and derived satellite-based intensity parameters are compared to severe weather reports. Young, growing thunderstorms are intensity rated using the rate of decrease of the satellite-observed cloud top minimum equivalent blackbody temperature (T_{BB}). Most storms, however, cannot be observed until they penetrate through the cirrus overcast produced by previous convection. Minimum cloud temperature and rate of T_{BB} isotherm expansion are then used as intensity indicators. In the following sections results from the case studies are presented and problems and limitations of the techniques and the data are discussed. In addition, an analysis of cloud top height variations (as viewed by the satellite) in relation to tornado touchdown times and, in a few cases, to mesocyclone formation times is also presented.

Table 1
Severe thunderstorm case study days

Date	Time (GMT)	Area	Spacecraft
May 6, 1975	06/1800 - 06/2200	S. Dakota to Texas	SMS-2
April 24, 1975	24/2000 - 25/0200	Oklahoma - Missouri	SMS-2
March 20, 1976	20/1800 - 20/2100	Missouri - Illinois	GOES-1
May 20, 1977	20/1700 - 21 0200	Oklahoma - Texas	GOES-1

Analysis of the satellite data is performed on the Atmospheric and Oceanic Information Processing System (AOIPS), an interactive image display and computer system.

2. OBSERVATIONS OF YOUNG, GROWING THUNDERSTORMS

When the satellite view of young, growing thunderstorms is not obscured by blowoff from nearby mature storms, the cloud top rate of rise in the 6-10 km layer can be estimated with the short-interval IR data. Figure 1 shows a composite of cases from May 6 and April 24, 1975. Two profiles are shown; one for cloud elements with associated severe weather reports (based on the National Severe Storms Forecast Center log), and one for storms with no reports. Especially in the 7-9 km layer there is a strong difference between the two mean profiles. The numbers in parentheses indicate the number of clouds averaged for each point. The two mean values in the 235-240 K layer, for example, are significantly different at the 1% level, using a t-distribution test. Thus, information on the intensity of the relatively young convection appears to be obtainable from the satellite data, and the time rate of change of minimum T_{BB} appears to be correlated with severe weather reports.

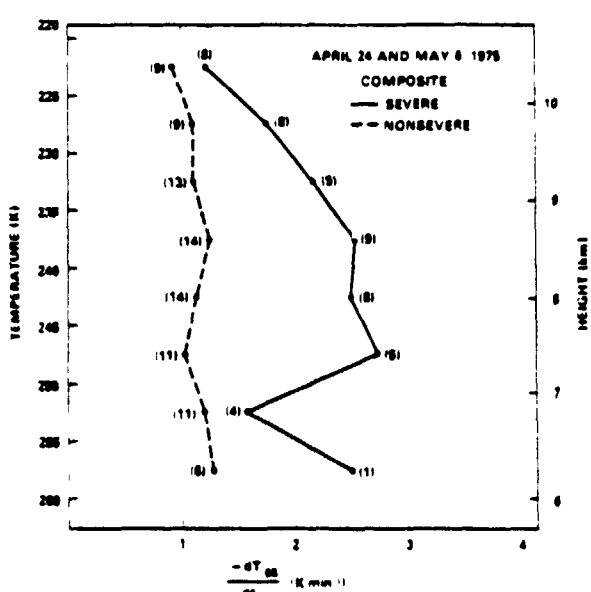


Figure 1 Composite profiles of cloud top ascent rates in the 6-10 km layer for severe and non-severe thunderstorms on May 6 and April 24, 1975.

The values of $-dT_{BB}/dt$ can be converted into vertical velocities using an appropriate lapse rate. The mean severe storm value of 2.5 K min^{-1} translates to an approximate value of 5 ms^{-1}. The interpretation of these cloud top ascent rates as vertical velocities and the comparison with other similar observations are discussed in detail in Adler and Fenn (1979b).

A representation of the thunderstorm features seen in the satellite IR is given in Figure 2 for a region of southern Oklahoma on April 24, 1975. Also shown are radar echo intensity contours from the National Weather Service radar at Oklahoma City. The storms are located approximately 90 km to the south and southeast of the radar site. On the satellite side of the diagram the cloud identification number and minimum T_{BB} are noted in each panel. The shaded blocks are the areas covered by the satellite data points with the minimum temperature. The smallest possible area is exemplified by the shaded area in cloud 22 at 2226 GMT. The additional outlines are T_{BB} isotherms at higher temperatures that are associated with that particular cloud element.

Figure 2 illustrates that the cloud elements being identified in the satellite data correspond to individual thunderstorms as evident in the radar presentation. Cloud 22 has a very rapid decrease of T_{BB} with time, and is associated with reports of large hail. Cloud 20 also has a hail report associated with it, while the other defined clouds in Figure 2 do not.

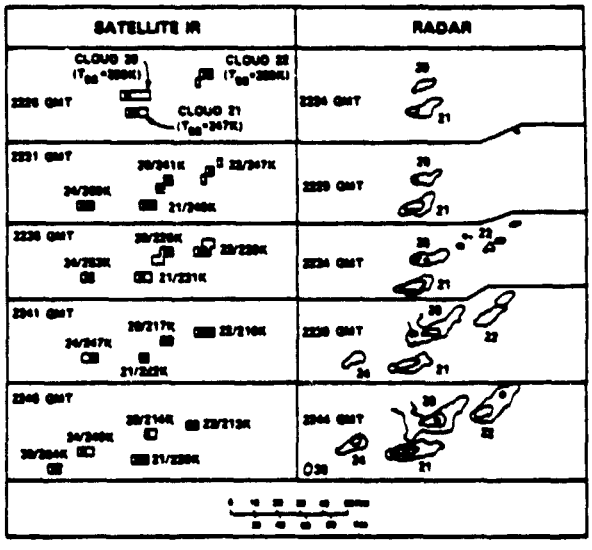


Figure 2. Evolution of April 24, 1975 thunderstorms in southern Oklahoma as observed by satellite IR and radar measurements.

3. ANALYSIS OF MATURE THUNDERSTORMS

As the growing thunderstorms penetrate the tropopause region they decelerate and the measured ascent rates become much lower. Figure 3 displays this effect in the 10–13 km region for a composite of storms on two days. Many storms cannot be viewed by the satellite until they penetrate up through middle and high cloud decks often produced by previous convection. New thunderstorms can sometimes only be identified when they reach very cold temperatures (<215 K). This lack of observability at lower heights (warmer temperatures) makes the use of $-dT_{BB}/dt$ as an intensity parameter more difficult. As we can see in Figure 3 the rate of temperature decrease is approximately 0.2 K min^{-1} at temperatures lower than 215 K. This means that a time interval of 5 minutes between images will result in only 1 K change. With the effective temperature resolution of the satellite data being about 1.5 K at this level the accuracy of calculated temperature change rates deteriorates.

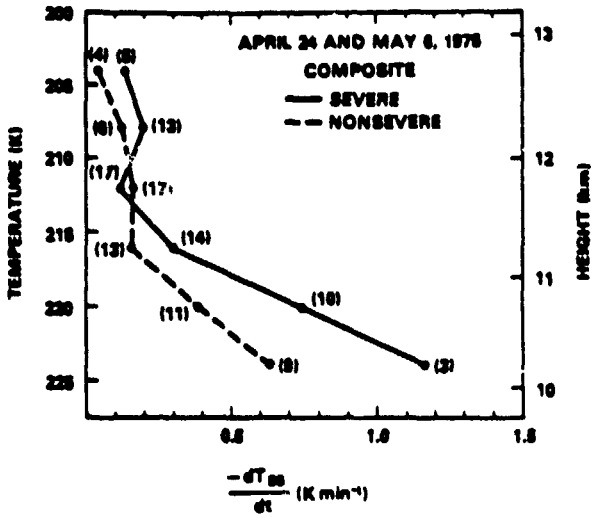


Figure 3. Composite profiles of cloud top ascent rates in the 10–13 km layer for severe and non-severe thunderstorms on May 6 and April 24, 1975.

Therefore, new intensity parameters for mature thunderstorms are necessary. Two parameters that have shown some skill are the minimum cloud top temperature itself and the normalized areal rate of expansion of T_{BB} isotherms. These variables have been discussed by Adler and Fenn (1979a) for one case study (May 6, 1975). Figure 4 (adapted from Adler and Fenn, 1979a) displays the distribution of 39 thunderstorms on May 6, 1975 along a cold front from South Dakota to Texas. The variables are the minimum T_{BB} achieved during the lifetime of the thunderstorm (a measure of maximum height) and the maximum $\frac{1}{N}\frac{dN}{dt}$ where N is the number of satellite data points with $T_{BB} \leq T_i$, where T_i is a particular cloud top T_{BB}. The expansion rate, $\frac{1}{N}\frac{dN}{dt}$, is interpolated between N=6 and N=30. Each data point represents an area of 22 km^2. Measuring the expansion rate between N equal to 6 and 30 results in a growth rate applicable to an individual thunderstorm. The isotherm rate of expansion is related to a combination of cloud top vertical velocity and divergence as discussed by Adler and Fenn (1979a).

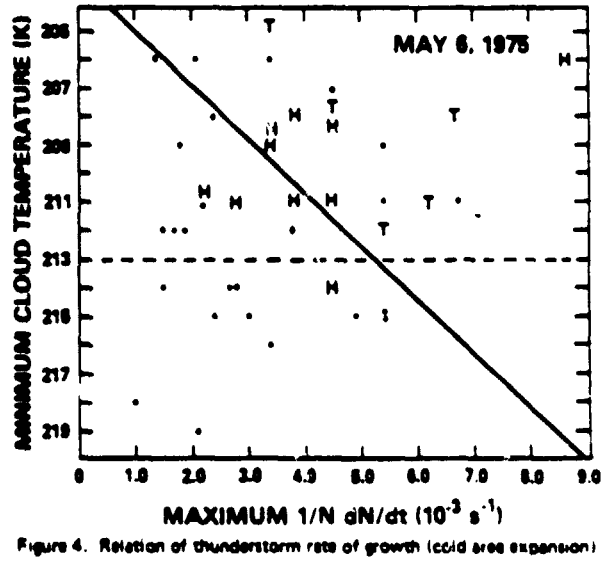


Figure 4. Relation of thunderstorm rate of growth (cold area expansion) to cloud top relative temperature and the occurrence of severe weather for 39 thunderstorm elements in analysis area. T indicates tornado. H represents reported hail. Horizontal dashed line represents tropopause temperature.

In Figure 4 the expansion rate parameter used to indicate the storm intensity is the maximum $\frac{1}{N}\frac{dN}{dt}$ at any $T_{BB} \leq 226$ K, although some of the storms were observed only at temperatures below 215 K. The diagram for this day does not show an obvious correlation between both T_{min} and $\frac{1}{N}\frac{dN}{dt}$ and severe weather, where H indicates hail, T represents tornado, and a dot means no accompanying severe weather reports.

The diagonal line in Figure 4 is a line of discrimination derived from the May 6, 1975 data. The horizontal dashed line is at the approximate tropopause temperature for the region. On this diagram it is a good temperature divider between severe and non-severe storms.

Results for a second case (April 24, 1975) in eastern Oklahoma and southwestern Missouri are shown in Figure 5. The diagonal line is the discriminant line derived from the May 6, 1975 case. The two parameters show an obvious correlation to the severe weather reports and, in this case, to each other. The tropopause temperature (also at 213 K for April 24) is not as good a discriminator between severe and non-severe storms as in the May 6 case.

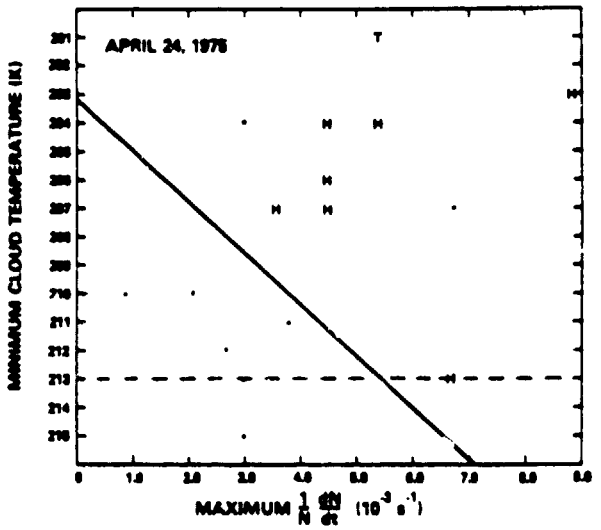


Figure 5. Same as Figure 4 except for 16 storms on April 24, 1975.

The next case (March 20, 1976) presents some additional complexities. The results are shown in Figures 6 and 7 for two adjacent areas. The "south area" is primarily in extreme eastern Missouri and southern Illinois. The "north area" is located in central Illinois. Although there is only a short distance between the two areas, the two regions are divided by the axis of the jet stream and have quite different tropopause temperatures and heights. This difference is evident on soundings (not presented) at Peoria, IL (north area) and Salem, IL (south area). Subjectively interpolating from the sounding locations to the storm locations, we selected 212 K and 209 K as representative tropopause temperatures for the north and south areas, respectively.

The minimum temperatures of the thunderstorms are on the average much colder in the south area (Figure 6). Also striking in the two diagrams is that all the defined storms have minimum temperatures at least 5 K below the tropopause temperature. This temperature difference was not evident in the two previous cases, and indicates the possibility of a difference in calibration among the cases. The March 20 case was analyzed with data from the GOES-1 spacecraft, while the other two cases were handled with SMS-2 data.

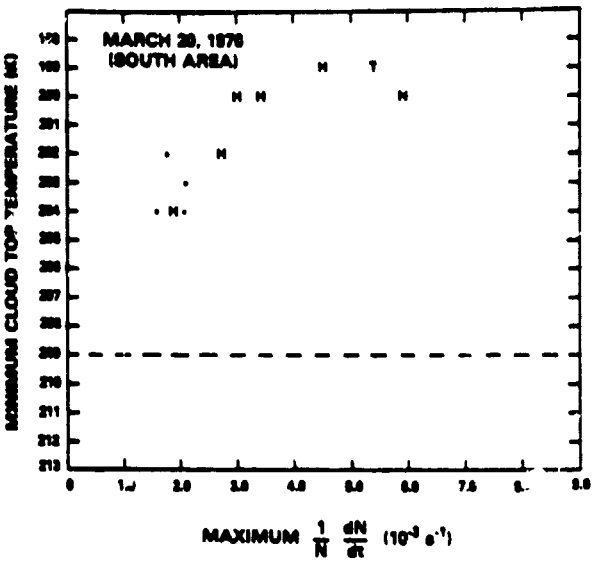


Figure 6. Same as Figure 4 except for 11 storms in the "south area" on March 20, 1976.

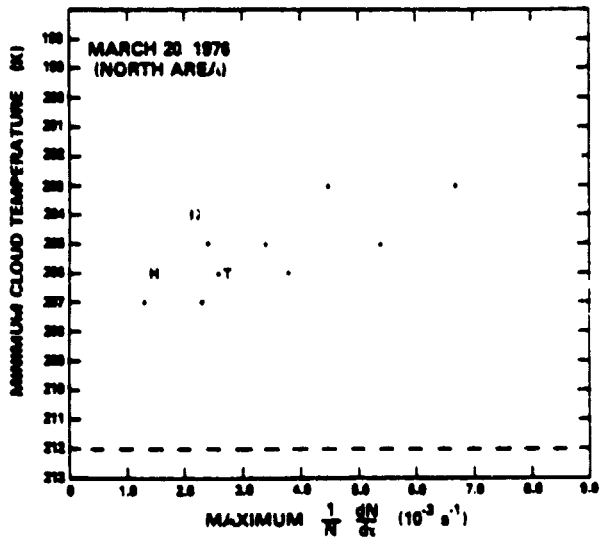


Figure 7. Same as Figure 4 except for 17 storms in the "north area" on March 20, 1976.

In the south area (Figure 6) the severe weather reports are grouped with the most intense storms and the two satellite-based parameters are correlated. For the north area (Figure 7), however, the three severe storms have relatively small rates of growth and average minimum temperatures.

The mature thunderstorms defined in the IR satellite data generally correspond to individual radar echoes. For example, Figure 8 shows a comparison of subjectively remapped contoured radar data (level 2 and above) and selected T_{BB} isotherms from the satellite IR data. Clouds 11, 10a, 10 and 7b had severe weather associated with them, although not necessarily at the time indicated. The correspondence between defined clouds in the satellite data and radar echoes is good, with the exception of cloud 7 which only appears on the radar as an elongated echo. A few weak echoes to the northwest of the line do not appear at the 226 K level in the satellite data.

The results of the May 6 and April 24, 1975 cases indicate that the minimum T_{BB} and the thunderstorm rate of growth

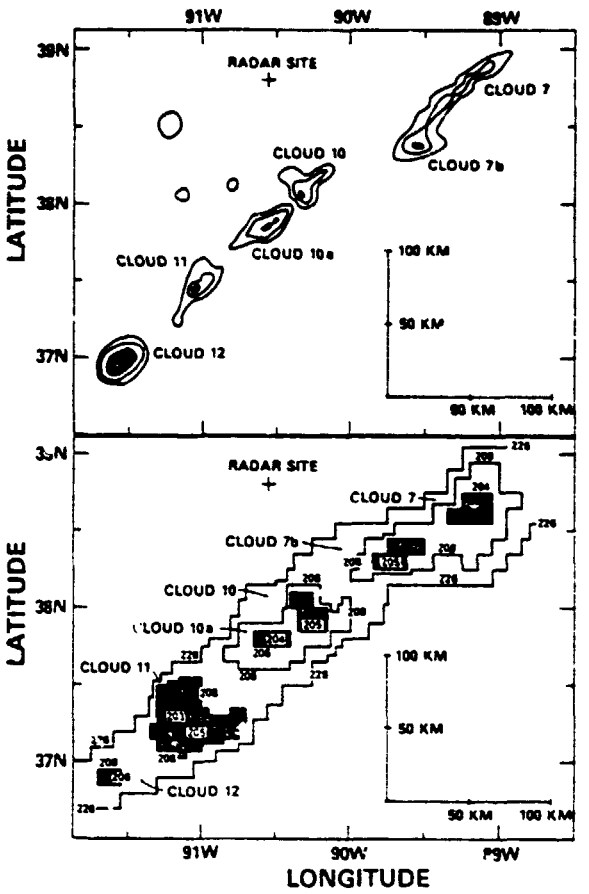


Figure 8. Distribution of radar echoes (from the St. Louis radar) and satellite IR features along a thunderstorm line across the Missouri-Illinois border on March 20, 1976 at 1903 GMT. Only level 2 and above radar intensity is shown.

parameter are good measures of storm intensity when compared to severe weather reports. Results from the March 20, 1976 case are not as clear but have revealed some additional complexities. As expected, variations in tropopause temperature must be considered. This is true both for comparing results between different days, but also for spatial variations over even relatively small areas on the same day. An additional complicating factor may be a varying calibration of the IR channel such that direct comparison of various days is difficult, especially when data from different spacecraft are used. A calibration variation of 3–4 K at the cold (200 K) end of the temperature range may result in ambiguous results when comparing different days.

Despite these problems the results so far show that even in the mature stage the satellite data can give valuable information on storm intensity and this intensity information is correlated with severe weather reports. Adler and Fenn (1979a) indicate a potential warning lead time of 30 minutes for these parameters.

4 CLOUD TOP CHARACTERISTICS OF TORNADIC STORMS

In this section the time variations of minimum T_{BB} of tornado-bearing thunderstorms on the four study days are examined. Five tornado bearing clouds on May 6, 1975 were discussed by Adler and Fenn (1979a). An examination of the five cloud elements having eight tornadoes clearly associated with them indicates that in seven of the eight cases the first report of the tornado took place during, or just after, a rapid expansion of cloud top cold areas (T_{BB} isotherms). This rapid expansion of areas within isotherms implies ascent. In this section these five storms will be reexamined along with six storms from the other three days studied.

Figure 9 displays minimum satellite-observed cloud top temperature (T_{BB}) as a function of time relative to tornado touchdown. The time of the tornado touchdown is noted by the vertical dashed line at time equal to zero. If there is more than one tornado associated with a cloud, only the most intense tornado is considered. Such is the case in panels a, c, d and f. In three cases (i, j, k) from May 20, 1977 the initial times of mesocyclones observed by the Doppler radar of the National Severe Storms Laboratory in Norman, Oklahoma are shown by the vertical dash-dot lines.

Eight of the 11 storms (a, b, c, d, e, i, j, k) have an associated rapid decrease of cloud top T_{BB} (as observed from satellite) approximately 30–45 minutes before tornado touchdown. This implies a rapid upward movement of the cloud top at this time. On most of these eight storms the rapid decrease in temperature is followed by a reduction in the slope of the temperature – time profile, and in the May 20 cases (i, j, k) a slight warming, indicating a cessation in upward growth or a slight drop in maximum cloud height.

This sequence of events at the cloud top can be pictured as being associated with storm evolution in the following way. A period of rapid upward growth at cloud top is associated with or possibly precedes the formation of the mesocyclone at mid-levels in the storm. Tornado touchdown is usually associated with a decreasing or constant maximum cloud top. This sequence of events at cloud top is consistent with radar studies (e.g. Lemon et al., 1978) and with hypotheses proposed by Fujita (1973). The satellite observations are on a scale of approximately 10 km, while the radar observations and aircraft observations of overshooting tops are on much smaller scales.

Three storms (f, g, and h) show decreasing cloud top T_{BB} (increasing height) at the time of the tornado touchdown. The tornadoes from all three of these storms were intense with Fujita/Pearson scale ratings of F4, F2, and F4 for the storms in panels f, g, and h, respectively.

Therefore, a majority of the storms analyzed have rapid ascent of cloud tops 30–45 minutes before tornado touchdown at apparently the time of the formation of the mesocyclone. A minority of the storms, however, have ascending cloud tops (at the scale of the satellite observation) at the time of tornado touchdown. The vertical velocities involved are small, but are reasonable when the horizontal scale (10 km) on which they are applicable is considered. A typical rate of temperature (T_{BB}) decrease in Figure 9 is 0.4 K min^{-1}. Assuming a lapse rate of 8 K km^{-1} the calculation of temperature change rate of 0.4 K min^{-1} converts to a vertical velocity of approximately 0.8 ms^{-1}.

It is interesting to note in Figure 9 that only panels i, j, k show a T_{BB} increase (indicating descent or cloud top collapse) prior to, or at the time of, tornado touchdown. The May 6 cases (a–e) show a constant temperature with time with some indication of a weak decrease at the time of the tornado. This difference may be related to the difference in larger scale cloud structure between the two cases and the satellite sensor's response characteristics. The May 6 storms appeared in the satellite images as a narrow line lying along a generally north-south oriented cold front. The east-west extent of the storms' anvils increased with time. As the satellite IR sensor scans from left to right across the scene it moves abruptly from a warm target (ground) to a cold target (thunderstorm top). With a narrow cirrus shield the sensor may not have enough time to accurately respond to the minimum T_{BB}. That is, for narrow (in the left-to-right direction) cirrus anvils we are probably overestimating the minimum T_{BB} (i.e., estimating it to be warmer than it is). As the narrow anvil expands with time the amplitude of the

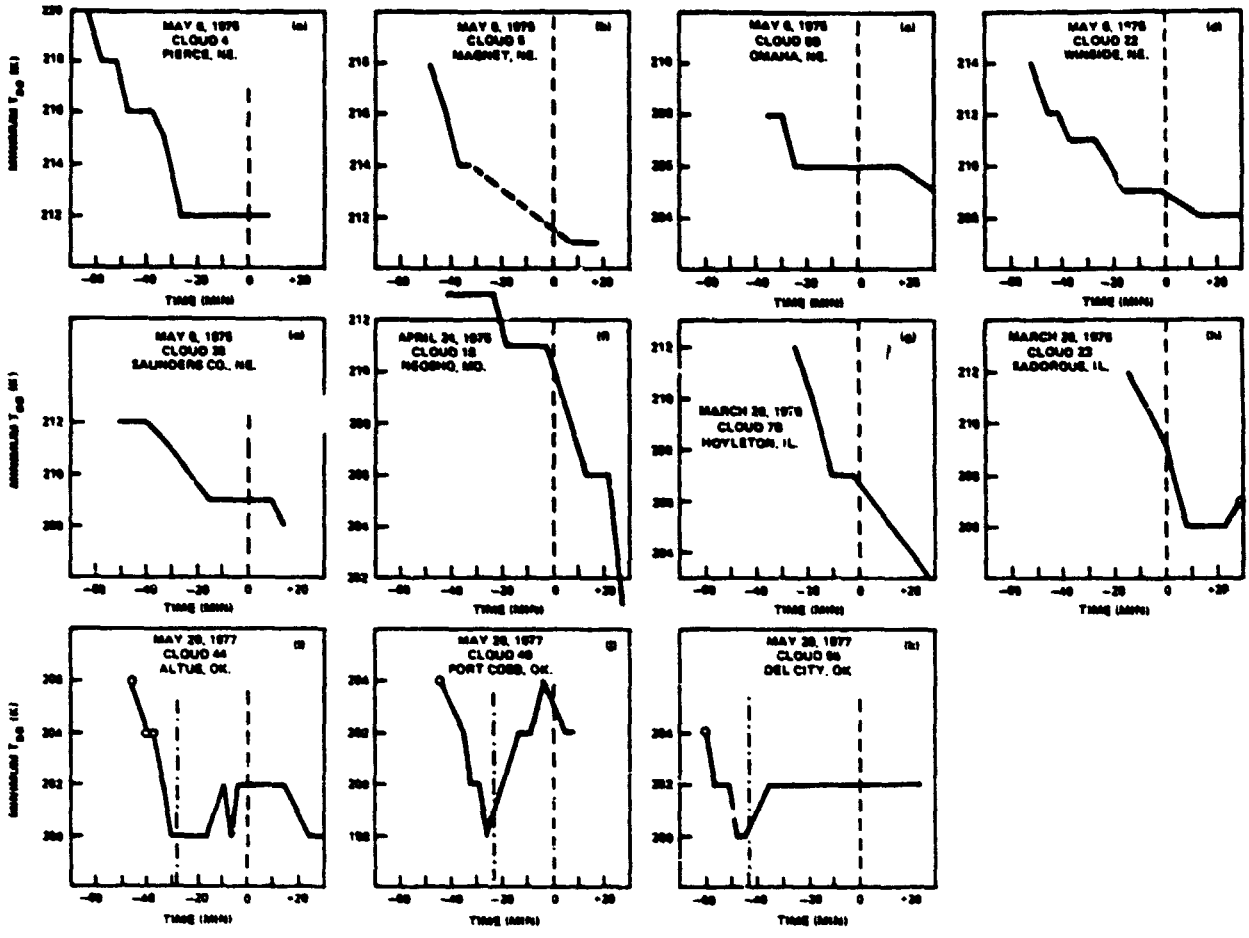


Figure 9. Satellite-observed cloud top temperature as a function of time relative to tornado touchdown times (vertical dashed line) and mesocyclone formation times, when available (panels i, j, k; vertical dash-dot line).

temperature overestimation should decrease as the distance from the left anvil edge to the storm center increases. If, while the anvil is expanding, the actual minimum T_{BB} is warming slightly, this could be obscured by the varying overestimation effect, with the result being a near constant observed T_{BB} with time.

On May 20, 1977 the storms (i, j, k) are imbedded in a very large cirrus shield with a large distance between the left edge and the storms' centers. Thus the sensor response problem in terms of estimating the minimum T_{BB} should not exist. In these three cases a warming is observed. Work on a technique to remove this sensor response effect is planned.

5. SUMMARY

Short-interval geosynchronous satellite data, in particular that from the IR channel, have the potential to be useful in the estimation of thunderstorm intensity. Since intensity is correlated with the occurrence of severe weather, such intensity information could be used for severe thunderstorm detection. However, there are some problems in the interpretation of the data, especially between different case study days involving different satellites.

The most direct measure of thunderstorm intensity can be made for young, growing thunderstorms. At a height of approximately 8 km there is a significant difference in cloud top ascent rates (as measured by $-dT_{BB}/dt$) between severe and non-severe thunderstorms. If intense thunderstorms can be detected at this altitude, there is an average potential warning lead time of about 60 minutes.

Most thunderstorms analyzed in this study could not be viewed at lower altitudes (such as 3 km) because of overlying cirrus due to previous convection. The satellite does not "see" the thunderstorm in the IR until it penetrates through these higher clouds. Often the thunderstorm top is then at a T_{BB} of 215K or even colder. At and above the region of the tropopause, the cloud top is decelerating and direct ascent rate calculations become inaccurate. As alternative measures of intensity we can use the minimum T_{BB} of the cloud (a relative measure of maximum height) and the normalized rate of expansion of areas within T_{BB} isotherms $\left(\frac{1}{N}\frac{dN}{dt}\right)$. In two of the case study days examined (May 6 and April 24, 1975) these parameters are correlated with severe weather reports. In the third case (March 20, 1976) the results are mixed. In one geographical area there is a good correlation, while in another area there is not. An examination of these three cases indicates that variations in tropopause temperature and probably satellite calibration must be taken into account before combining the data sets from different days.

Eleven cases of tornadic thunderstorms were examined with respect to cloud top temperature (height) variations relative to tornado touchdown times, and in three cases relative to the initial observation of mesocyclones by Doppler radar. In 8 of the 11 cases there is a period of rapid ascent 30–45 minutes prior to tornado touchdown. This upward growth appears to be associated with the formation of the mesocyclone. This ascent is followed by a period of no growth or even a drop in cloud top height preceding, or at the time of, tornado touchdown. In the three remaining cases cloud top ascent is evident in the satellite data at tornado touchdown.

REFERENCES

Adler, R. F. and D. D. Fenn, 1979a: Thunderstorm intensity as determined from satellite data. *J. Appl. Meteor.*, 18, 502-517.

Adler, R. F. and D. D. Fenn, 1979b: Thunderstorm vertical velocities estimated from satellite data. To appear in *J. Atmos. Sci.*

Fujita, T. T., 1973: Proposed mechanism of tornado formation from rotating thunderstorms. In *Preprints of Eighth Conference on Severe Local Storms.*

Lemon, L. R., D. W. Burgess, and R. A. Brown, 1978: Tornadic storm outflow and morphology derived from single-Doppler radar measurements. *Mon. Wea. Rev.*, 106, 48-61.

Paper 37

Reprinted from JOURNAL OF THE ATMOSPHERIC SCIENCES, Vol. 36, No. 9, September 1979
American Meteorological Society
Printed in U. S. A.

Thunderstorm Vertical Velocities Estimated from Satellite Data

ROBERT F. ADLER

Laboratory for Atmospheric Sciences, Goddard Space Flight Center, NASA, Greenbelt, MD 20771

DOUGLAS D. FENN

GE/MATSCO, Beltsville, MD 20705

(Manuscript received 9 January 1979, in final form 3 May 1979)

ABSTRACT

Infrared geosynchronous satellite data with an interval of 5 min between images are used to estimate thunderstorm top ascent rates on two case study days. A mean vertical velocity of 3.4 m s^{-1} for 23 clouds is calculated at a height of 8.7 km. This upward motion is representative of an area of approximately 10 km on a side. Thunderstorm mass flux of $\sim 2 \times 10^8$ kg s^{-1} is calculated, which compares favorably with previous estimates. There is a significant difference in the mean calculated vertical velocity between elements associated with severe weather reports ($\bar{w}$ = 4.9 m s^{-1}) and those with no such reports (2.4 m s^{-1}).

Calculations were made using a velocity profile for an axially symmetric jet to estimate the peak updraft velocity. For the largest observed w value of 7.8 m s^{-1} the calculation indicates a peak updraft of ~50 m s^{-1}.

1. Introduction

Vertical velocity is a basic parameter in meteorological analysis and forecasting. However, unlike horizontal velocity, it usually cannot be observed directly. The presence of upward air motion can often be inferred by the occurrence of clouds and precipitation, but quantitative evaluation of vertical motion must usually be made through the omega equation, the adiabatic method or a similar indirect technique. An exception to this is the recent work with multiple-Doppler radar data.

In the current paper we present the results of using a simple method to estimate thunderstorm cloud top vertical velocity from SMS/GOES rapid-scan (5 min interval) window channel infrared (IR) data. Time rate of change of cloud-top minimum equivalent blackbody temperature T_{BB} is converted to vertical velocity w by

$$w = \left(\frac{\partial T}{\partial z}\right)^{-1} \frac{dT_{BB}}{dt}, \qquad (1)$$

where the lapse rate is determined from rawinsonde data. An example calculation using SMS data and Eq. (1) was given by Adler and Fenn (1976). In the following sections the calculated vertical velocities are compared for clouds with associated severe weather reports and for those with no such reports, and the computed values are also compared with previous estimates of thunderstorm vertical velocities.

2. Description of analysis technique

The analysis of the digital satellite data was performed on the Atmospheric and Oceanic Information Processing System (AOIPS), an interactive image analysis system described by Billingsley (1976). Sequences of images are enhanced, thunderstorm elements are isolated and identified by locating relative minima of T_{BB} in the IR images. Their maximum gray level (minimum T_{BB}) are then recorded. The analysis of the thunderstorms used in this study (on 24 April 1975 and 6 May 1975) was part of a larger effort (see Adler and Fenn, 1979). On each of the study days an area of convection was monitored during a set period of time (~4 h). Thunderstorm elements were defined and time histories of each element were determined.

The emphasis in the earlier work (Adler and Fenn, 1979) was on convective elements after they had reached ~10 km in height (as estimated by the satellite IR data). This paper highlights observations of thunderstorms earlier in their growth cycle as they penetrate upward through the middle troposphere. It should be empnasized that not all thunderstorms can be observed at middle tropospheric heights. This is because they are often hidden by dense cirrus clouds produced by previous convection. In the two case studies that will be described, of the thunderstorm elements defined above 10 km (T_{BB} = 226 K), only about 25–30% could be detected at lower heights. On typical con-

0022-4928/79/091747-08$06.00

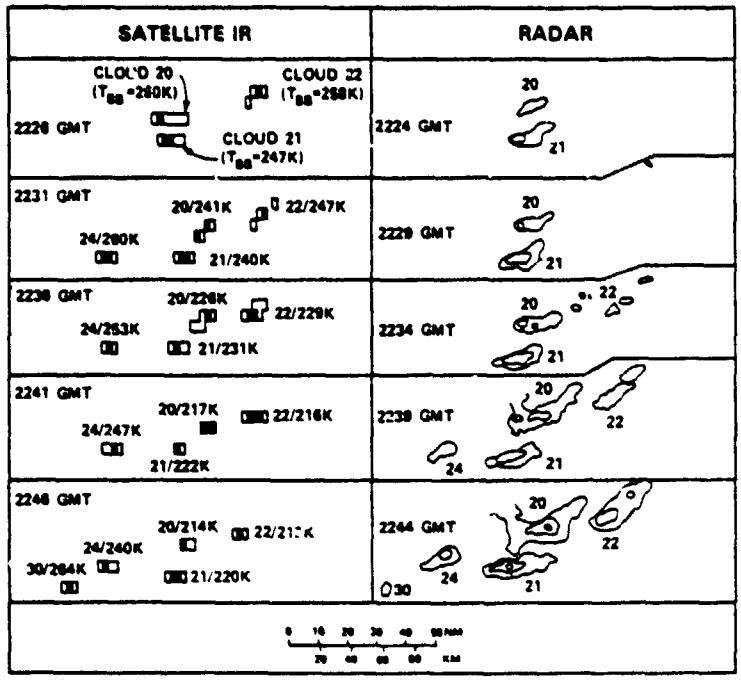


FIG. 1. Evolution of 24 April 1975 thunderstorms in southern Oklahoma as observed by satellite IR and radar measurements.

vection days this percentage will vary with time, being large early in the day as convection begins and decreasing as the buildup of thunderstorm-produced high clouds obscures younger convective elements. The results described in this paper concern that relatively small group of thunderstorms that could be identified at lower levels.

A representation of the thunderstorm features seen in the satellite IR is given in Fig. 1 for a region of southern Oklahoma on 24 April 1975. Also shown are radar echo intensity contours from the National Weather Service radar at Oklahoma City. The storms are located ~90 km to the south and southeast of the radar site. On the satellite side of the diagram the cloud identification number and minimum T_{BB} are noted in each panel. The shaded blocks are the areas covered by the satellite data points with the minimum temperature. The smallest possible area is exemplified by the shaded area in cloud 22 at 2226 GMT. The additional outlines are T_{BB} isotherms at higher temperatures that are associated with that particular cloud element.

Fig. 1 illustrates that the cloud elements being identified in the satellite data correspond to individual thunderstorms as evident in the radar presentation. Cloud 22 has a very rapid decrease of T_{BB} with time, and is associated with reports of large hail. Cloud 20 also has a hail report associated with it, while the other defined clouds in Fig. 1 do not.

3. Vertical velocity estimates

a. Sources of error

The vertical velocity estimates presented in this paper are subject to error because of uncertainties in the satellite radiance measurements and possible unrepresentativeness of the data. The first question is: How are cloud top ascent rate and vertical velocity related? Even if the satellite measurements are without errors and are representative, the w calculated using Eq. (1) is actually the cloud top ascent rate. Due to mixing at the cloud top this may not be identical to the vertical air motion at the cloud top. For example, Woodward (1959) shows laboratory results indicating that the center of an isolated thermal rises at approximately twice the speed of the thermal cap. However, this is the case for a small, decelerating thermal, with extensive mixing through the sides. In addition, for a decelerating thermal the center of the thermal, with a vertical velocity of w_1 at time t_1, will show a decrease of w with time so that when (at t_2) it reaches an altitude equal to that of the thermal cap at the earlier time t_1 its velocity will have decreased significantly to w_2, much closer in magnitude to the rate of rise of the thermal cap as it passed that altitude. Thus for the scale of observation in the current analysis, the cloud top ascent rate is probably nearly equal to the magnitude of the vertical air flow at that level.

Another possible source of error is in the cloud

emissivity. For the thick water clouds that were part of this study, the emissivity (ϵ) is very close to unity. In order to use T_{BB} in place of T in Eq. (1), $\epsilon = 1$ must be assumed. For ice clouds, especially thin cirrus clouds, the emissivity can be substantially less than 1.0. As the thunderstorm top glaciates, the change from water to ice may reduce the cloud top emissivity and T_{BB} will be larger than T, the cloud-top temperature. This effect, however, appears to be small and will not affect the vertical velocity calculations. This conclusion is based on calculations made with the help of tables presented by Hunt (1973). For a water cloud with liquid water content (LWC) of 4×10^{-4} kg m^{-3}, 95% of the emitted radiation at the 11 μm wavelength comes from the top 84 m of the cloud. For an ice cloud with the same LWC value, the depth of extinction (to the 95% level) is 150 m. Therefore, the water-ice variation in cloud-top structure will cause only a very small (<1 K) variation in T_{BB}. Calculations based on coefficients presented by Cox (1977) also support this conclusion.

The most serious source of error stems from the satellite data itself. The IR channel has an instantaneous field of view (IFOV) of 8 km on a side at the subsatellite point and approximately 10 km at 40°N. As noted by Negri *et al.* (1976) the use of SMS/GOES IR temperatures to determine thunderstorm height results in underestimates, especially for small elements. In comparison with radar measurement of thunderstorm tops, the satellite underestimation is ~2 km. This effect is related to two factors. First, the satellite, because of its rather large IFOV, is averaging over an area of ~100 km^2 compared with the radar observation, which is applicable to a much smaller area. Thus the radar will be identifying smaller, higher features because of its better resolution. The second factor is inadequate sensor response when going from a warm (low) to a cold (high) target (Negri *et al.*, 1976). In general, the bias in the estimation of storm height will not significantly affect the vertical velocity calculation. However, it may affect the height to which the velocities are assigned.

Other errors might arise from the use of inaccurate lapse rates in Eq. (1). In the calculations to follow we use a smooth profile which is a mean of the ambient (determined from rawinsonde data) and the moist adiabatic lapse rate. This compromise was chosen because it was thought that, although the air in updraft cores of large thunderstorms is believed to ascend moist adiabatically, the large area represented by the satellite data implies a large environmental effect. The calculations are also rather insensitive to variations in the lapse rate. Using either the moist adiabatic or the ambient lapse rate instead of the average of the two produces only 10% differences in the calculated velocities.

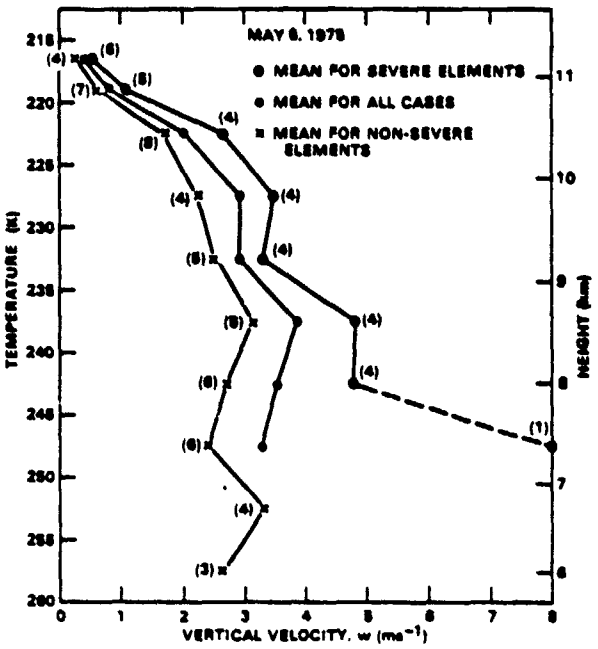


FIG. 2. Composite vertical velocity profiles for 6 May, 1975. Numbers in parentheses are the number of cases.

The final validation of the calculated vertical velocities must come through a careful comparison with radar, aircraft or satellite stereo observations. Some of the work toward these comparisons is underway. Based on the previous discussion of errors, however, the authors believe the values calculated from (1) are accurate to within 20%, for the size area they represent. Comparison with other estimates and calculations will be made in the following sections.

b. Vertical velocity results

Fourteen elements on 6 May 1975 and 14 on 24 April 1975 were analyzed. Minimum T_{BB} as a function of time for each element was plotted and dT_{BB}/dt values were calculated. The warmest or lowest of the monitored elements were at T_{BB} = 260 K (~6 km). Not all thunderstorms could be observed from that point upward through the remainder of the troposphere. Some elements were obscured by other storms; other elements were not detected until they penetrated middle level cloud fields.

Vertical velocities were calculated every 5 K in the vertical for each cloud element or thunderstorm using Eq. (1) and a lapse rate halfway between ambient and moist adiabatic. The mean w was then calculated for various categories of clouds to produce composite profiles. The results for the 6 May case are shown in Fig. 2. Profiles are shown for thunderstorms associated with severe weather reports (based on National Severe Storms Forecast

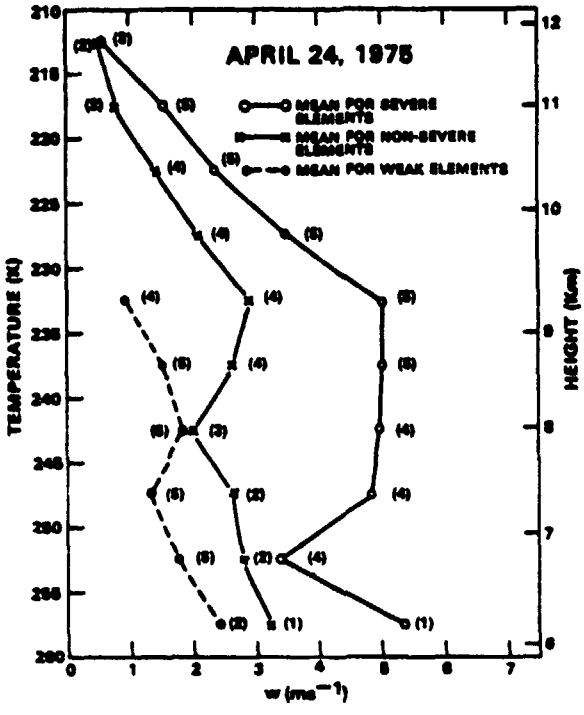


FIG. 3. Composite vertical velocity profiles for 24 April, 1975.

Center logs) and those with no accompanying reports. The numbers in parentheses indicate the number of cases constituting each mean or composite vertical velocity. The mean profile of all cases is also shown. All the elements defined in the 6 May case were intense thunderstorms reaching heights >10.5 km as determined from the SMS/GOES T_{BB}'s. These heights are probably low compared to radar echo top heights. The 6 May storms were located in a relatively narrow band from eastern Nebraska to Texas along a north–south oriented cold front.

The profile in Fig. 2 for all of the 6 May cases indicates a mean magnitude for w of -3.5 m s^{-1} at 8 km. Above that level there is a gradual decrease in the upward flow and a more rapid decrease above 10 km. Between 10 and 11 km the horizontal divergence can be estimated through the continuity equation, assuming incompressibility, so that

$$-\frac{\partial w}{\partial z} = \nabla \cdot \mathbf{V} = 2 \times 10^{-3}\ \mathrm{s}^{-1}. \qquad (2)$$

The vertical velocities in Fig. 2 and the calculated divergence noted above are applicable to an area of about 10 km on a side. The calculated vertical velocities *do not* represent updraft core velocities, which could be an order of magnitude larger when measured on a horizontal scale of 1 km (see Section 5).

The results displayed in Fig. 2 indicate that on the average the elements with associated severe weather reports have larger vertical velocities. This is not surprising since intensity of convection has always been closely associated with severe weather.

A similar diagram for 24 April 1975 is given in Fig. 3. The convection of interest on this day was centered in southwestern Missouri. The composite w profiles for three categories are displayed. The additional category is for weak elements which did not reach a height of 10 km (as determined by the T_{BB} values). In the layer from 7 to 9 km the average w is -1.5 m s^{-1} for those storms. This is significantly lower than the composite for "non-severe" elements in either Figs. 2 or 3. Part of the difference in w values may be due to a size difference and therefore a difference in the fraction of the IFOV filled by the ascending cloud. The severe elements have a mean w of 4–5 m s^{-1} in the 6–9 km height range, similar to that found in the 6 May case (Fig. 2).

The 235–240 K level (~8.7 km) is representative of the layer of relatively large vertical velocities on both days. Fig. 4 shows the frequency distribution of the 23 elements or clouds for these two days. The hatched portion of the histogram contains the values for the severe weather elements. The average w for all 23 cases is 3.4 m s^{-1}. The severe and non-severe elements have average values of 4.9 and 2.4 m s^{-1}, respectively. The severe thunderstorms dominate the high end of the distribution where six out of seven cases with $w > 4$ m s^{-1} are associated with severe weather reports. The difference in mean w between severe and non-severe clouds is significant at the 1% level using the t-distribution test (Panofsky and Brier, 1963).

c. Mass flux calculation

Because the values of w calculated in the last section are representative of an area larger than a typical thunderstorm updraft, vertical volume or mass flux calculations can be simply made from

$$F_m = \rho A w, \qquad (3)$$

where F_m is the vertical mass flux, ρ the density and A the area. For the 235–240 K layer (~8.7 km) ρ is assumed to be 5×10^{-1} kg m^{-3}, and A is assigned a value of 100 km^2 for the area of the satellite IFOV.

With the given values for ρ and A the mean w for all storms of 3.4 m s^{-1} is converted to a mass flux of 1.7×10^8 kg s^{-1}. The mean w of severe elements (4.9 m s^{-1}) is equivalent to a mass flux of 2.4×10^8 kg s^{-1}. These magnitudes are for the mass flux through a given layer associated with a growing thunderstorm top. The calculated values compare favorably with results presented by other investigators. Kropfli and Miller (1976) calculate a

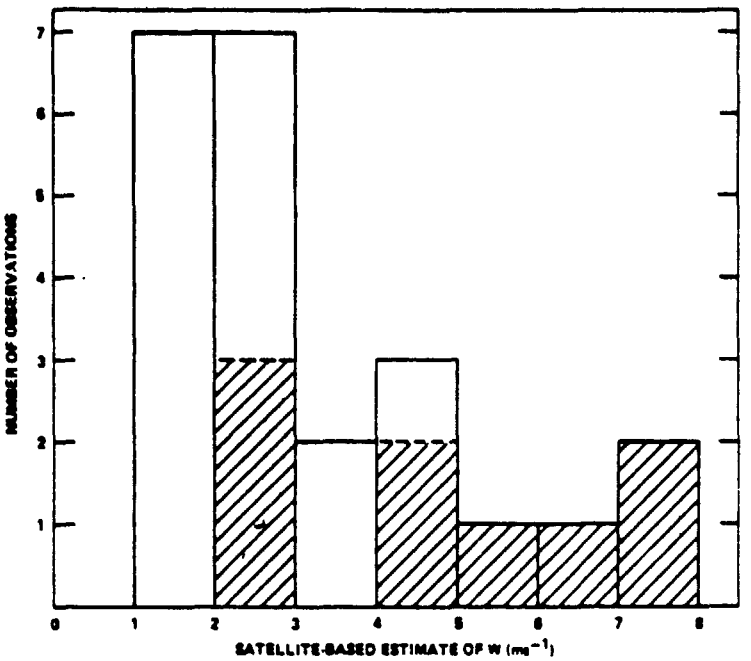


FIG. 4. Frequency distributions of estimated vertical velocities at 8.7 km for clouds on 24 April and 6 May, 1975. Hatched portion of histogram indicates thunderstorms with accompanying severe weather reports.

value of 1.9–2.0 × 10^8 kg s^{-1} between 8 and 9 km for a northeast Colorado storm calculated using vertical velocity inferred from dual-Doppler radar data. Auer and Marwitz (1968) present results of the cloud-base mass flux into 18 hailstorms on the high plains deduced from aircraft measurements. Their average value is 2.3 × 10^8 kg s^{-1}. Therefore, it appears that the upward flow deduced from the satellite observations is of a reasonable magnitude when compared to calculations and observations on approximately the same scale. Inferences about the magnitude of the maximum updraft are presented in Section 5.

4. Example of intense thunderstorm

On 24 April 1975 a severe thunderstorm complex developed over extreme northeastern Oklahoma in the late afternoon and moved into southwestern Missouri around sunset. The most significant severe weather associated with the system was the Neosho, Missouri, tornado which touched down at approximately 0040 GMT 25 April. By following the evolution of the storm system backward in time, the initial intense convection can be detected and its associated rapid cloud-top growth calculated.

The Neosho cloud system was designated cloud 18 as part of a larger study of this day. Fig. 5 exhibits minimum cloud-top T_{BB} as a function of time for cloud 18 in its early stages. The temperature drops precipitously between 2200 and 2220 GMT with a maximum calculated rate of 4 K min^{-1}. The drop in temperature between 260 and 220 K takes only a little more than 15 min. This type of rapid change emphasizes the importance of short-interval data to study and monitor thunderstorm activity.

Using a lapse rate varying from 7.8 K km^{-1} at 260 K, to 8.6 K km^{-1} at 240 K, to 8.0 K km^{-1} at 215 K the vertical velocity was calculated using Eq. (1). The results for cloud 18 are shown in Fig. 6. The maximum w is 7.8 m s^{-1} at about 9 km. Above that height there is a rapid decrease of w with height, with a calculated divergence [using Eq. (2)] of 4.2 × 10^{-3} s^{-1} over nearly a 2 km deep layer. It must be remembered that the vertical velocities and divergences are applicable for an area of ~10 km on a side.

The temperature curve in Fig. 5 flattens out and reaches a plateau at 214 K, which is equivalent to a height of ~11.5 km. Figs. 5 and 6 are indicative of the convective surge of the first thunderstorm cell in the system. After 2320 GMT the organization of the system (as viewed by the satellite) becomes more complex, with three cold centers appearing. The main center (designated cloud 18c) undergoes some short-lived growth associated with the Blue Jacket, Oklahoma, tornado, but then remains quiet, although definable, until a new cell evidently penetrates the already cold cirrus canopy. With this new cell the T_{BB} values again fall, this time to 206 K. During this period of decreasing T_{BB} values, the Neosho, Missouri, tornado touched down. A further

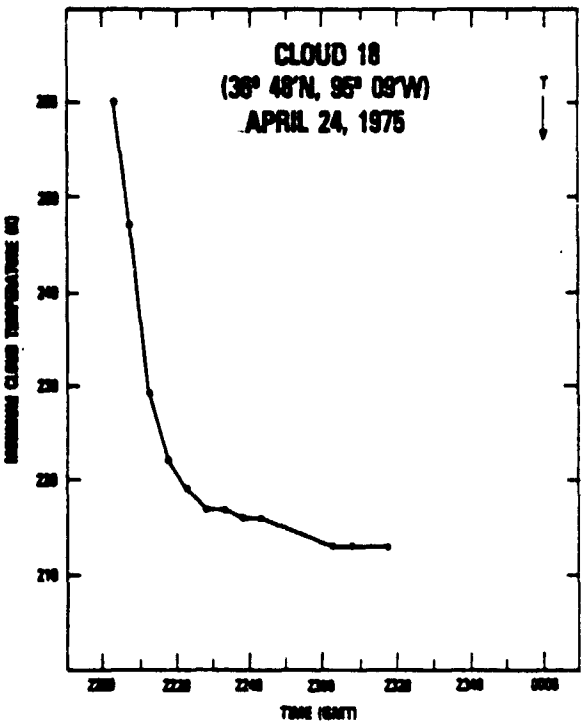


FIG. 5. Minimum equivalent blackbody temperature T_{BB} as a function of time for cloud 18 on 24 April 1975.

drop in temperature to 201 K occurs 20 min later. The minimum temperature on the nearby Monett, Missouri sounding was 210 K, indicating that the storm had penetrated well into the stratosphere. Although the initial disturbance (Fig. 5) leveled out at 214 K, it was the starting point for convection that eventually reached 201 K, about 13 km in height, 1 km above the tropopause.

5. Interpretation of calculated vertical velocities in terms of maximum updraft

The vertical velocities presented in the previous sections of this paper are mean velocities over an area equivalent to the satellite instantaneous field of view (IFOV), which in this case is ~100 km². In the temperature range 235–240 K (~8.7 km) the 23 observed w's based on the satellite data ranged from 1.2 to 7.8 m s^{-1}, with a mean of 3.4 m s^{-1}. Although these are very large values when compared to typical synoptic-scale w's, they are small when compared to maximum thunderstorm updraft magnitudes. Thunderstorm updrafts can reach magnitudes of 10 m s^{-1} very easily and are typically 30 m s^{-1} in supercell thunderstorms (Browning, 1977; Davies-Jones, 1974). These large updraft values are probably representative of an area ~1 km². Thus there are two orders of magnitude difference in the area covered by the estimated w's obtained from the satellite data in this study and the area covered by the peak updraft velocity.

Assuming axial symmetry and a knowledge of the shape and size of the radial profile of vertical velocity, one can make an estimate of the maximum updraft magnitude. Kyle *et al.* (1976) have investigated updraft profiles determined from penetrating aircraft and have fitted mathematical expressions to the observations. One of the expressions tested, with good results, is the profile for an axially symmetric jet (Schlichting, 1968). The formula is

$$w = w_0 e^{-a(r/R)^2}, \tag{4}$$

where w_0 is the peak w, r is the radial distance, R the radius of the updraft, and the constant $a = 2.3$. The constant was chosen in the present work so that $w = 0.1\, w_0$ at $r = R$. That is, the updraft radius R is defined so that the vertical velocity is not zero at the updraft edge, but one-tenth the maximum value.

Integrating Eq. (4) over a circular area of radius r_1 and dividing the result by the area of the integration produces an expression for the mean w over the area, i.e.,

$$\bar{w} = \frac{w_0}{a}\left(\frac{R}{r_1}\right)^2 [1 - e^{-a(r_1/R)^2}]. \tag{5}$$

For values of $r_1 > R$, the term in brackets approaches a value of 1.

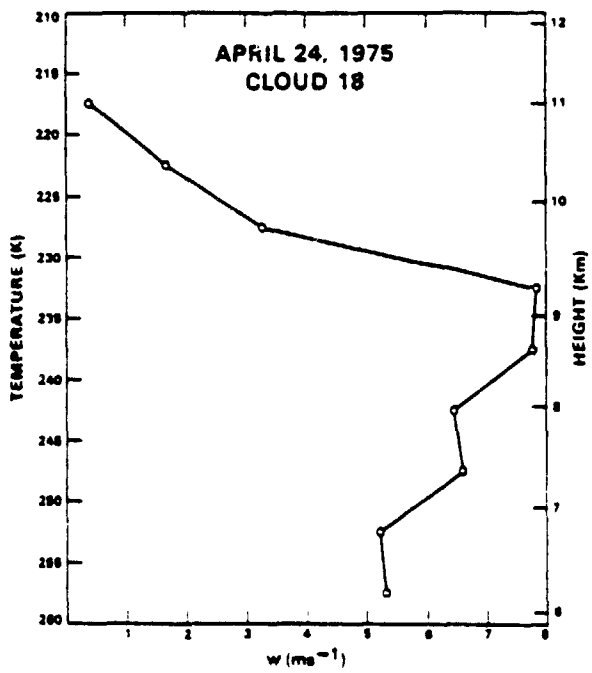


FIG. 6. Estimated vertical velocity profile for cloud 18.

If the radius of integration is equal to the updraft radius ($r_1 = R$),

$$\bar{w} = 0.39 w_0. \quad (6)$$

Therefore, if the updraft, as defined by R, exactly fills the satellite IFOV, the maximum updraft will still be 2.5 times the satellite-based estimate of $\bar{w}$.

The size of thunderstorm updrafts is highly variable (Browning, 1977). Based on the discussion of updraft sizes by Browning (1977) and an examination of cross sections by Kropfli and Miller (1976) and Ray (1976) of thunderstorm vertical motions deduced from Doppler radar data, an updraft radius of 3 km is reasonable.

For an area of 100 km^2 (approximately equal to the IFOV), the equivalent radius of integration is 5.6 km. Thus with $R = 3$ km and $r_1 = 5.6$ km,

$$\bar{w} = 0.12 w_0. \quad (7)$$

Thus, with all the assumptions as to profile shape and updraft size, Eq. (7) indicates that the mean $\bar{w}$ of 3.4 m s^{-1} is equivalent to a w_0 of 28 m s^{-1} and the 7.8 m s^{-1} value from the Neosho storm (Section 4) is equivalent to a w_0 of 65 m s^{-1}.

If the updraft radius is larger than 3 km, the w_0 values would be smaller than just calculated. For an R of 3.5 km, instead of 3 km, the 65 m s^{-1} value in the last paragraph would drop to 46 m s^{-1}. It is obvious that the calculated w_0 is sensitive to the size of the updraft. Observations summarized by Kyle *et al.* (1976) indicate that R is a function of w_0, at least for small radii. An accurate expression for R as a function of w_0 would allow Eq. (5) to become a statement of the relation between $\bar{w}$ and w_0.

Despite the variability and sensitivity of the w_0 calculations, it is evident that the satellite observations on a scale of 10 km are producing $\bar{w}$ of up to approximately 8 m s^{-1}, and this can be interpreted as being roughly equivalent to a maximum updraft of 50 m s^{-1} in intense thunderstorms.

6. Summary

Rapid-scan (5 min interval) SMS/GOES IR data have been used to estimate thunderstorm top ascent rates for severe and non-severe thunderstorms on two case study days. The vertical velocities are calculated by converting the time rate of change of minimum T_{BB} to vertical velocity w through use of a lapse rate. On both days examined (6 May and 24 April 1975) the thunderstorm elements with associated severe weather reports have larger average w's. At a temperature level (235–240 K) equivalent to a height of ~8.7 km, the 23 elements monitored had a mean w of 3.4 m s^{-1}. The severe and non-severe elements had mean w's of 4.9 and 2.4 m s^{-1}. Intensity of convection appears to be correlated with the occurrence of severe weather, and the satellite data appear to be capable of quantifying the convection intensity.

The calculated vertical velocities are representative of an area (100 km^2) roughly equivalent to the satellite instantaneous field of view (IFOV). Mass flux estimates of $\sim 2 \times 10^8$ kg s^{-1} are calculated, which are reasonable in comparison with other estimates.

The largest w, calculated directly from the satellite data (at the 8.7 km height), was 7.8 m s^{-1} occurring with the initial convection associated eventually with the Neosho, Missouri, tornado. This vertical velocity corresponds to a rate of cloud top temperature decrease of 4 K min^{-1}, which indicates the need for high time resolution geosynchronous satellite data even with the relatively coarse spatial resolution of the IR measurements.

Calculations were performed to estimate the peak updraft velocity from the satellite-based values (averages over 100 km^2 areas). Using the velocity profile formula for an axially symmetric jet to represent the thunderstorm updraft, we derived an expression relating $\bar{w}$ (the satellite-based w) to w_0 (peak updraft velocity) and R (updraft radius). With a reasonable value of R (3–3.5 km), the $\bar{w}$ of 7.8 m s^{-1} for the Neosho storm produces an estimate of approximately 50 m s^{-1} for w_0.

The calculations and examples of this paper reveal that the SMS/GOES rapid-scan data can provide useful quantitative information on thunderstorm vertical motions and therefore convection intensity. Although the vertical motion calculations must undergo closer scrutiny through comparison with Doppler radar estimates and possibly with vertical growth rates determined from stereo geosynchronous satellite images, they provide a starting point for the quantitative use of the satellite data in the study of thunderstorms and provide the possibility for eventual use in the monitoring of thunderstorm activity.

REFERENCES

Adler, R. F., and D. D. Fenn, 1976: Thunderstorm monitoring from a geosynchronous satellite. *Preprints 7th Conf. Aerospace and Aeronautical Meteorology*, Melbourne, Amer. Meteor. Soc., 307–311.

——, and ——, 1979: Thunderstorm intensity as determined from satellite data. *J. Appl. Meteor.*, **18**, 502–517.

Auer, A. H., and J. D. Marwitz, 1968: Estimates of air and moisture flux into hailstorms on the high plains. *J. Appl. Meteor.*, **7**, 196–198.

Billingsley, J. B., 1976: Interactive image processing for meteorological applications at NASA/Goddard Space Flight Center. *Preprints 7th Conf. Aerospace and Aeronautical Meteorology*, Melbourne, Amer. Meteor. Soc., 268–275.

Browning, K. A., 1977: The structure and mechanisms of hailstorms. *Hail: A Review of Hail Science and Hail Suppression*, *Meteor. Monog.*, No. 38, Amer. Meteor. Soc., 1–43.

Cox, S. K., 1977: Satellite derived cloud cluster cloud-top heights. *Preprints 11th Tech. Conf. Hurricanes and Tropical Meteorology*, Miami Beach, Amer. Meteor. Soc., 355–356.

Davies-Jones, R. P., 1974: Discussion of measurements inside high-speed thunderstorm updrafts. *J. Appl. Meteor.*, 13, 710–717.

Hunt, G. E., 1973: Radiative properties of terrestrial clouds at visible and infrared thermal window wavelengths. *Quart. J. Roy. Meteor. Soc.*, 99, 346–369.

Kropfli, R. A., and L. J. Miller, 1976: Kinematic structure and flux quantities in a convective storm from dual-Doppler radar observations. *J. Atmos. Sci.*, 33, 520–529.

Kyle, T. G., W. R. Sand and D. J. Musil, 1976: Fitting measurements of thunderstorm updraft profiles to model profiles. *Mon. Wea. Rev.*, 104, 611–617.

Negri, A. J., D. W. Reynolds and R. A. Maddox, 1976: Measurements of cumulunimbus clouds using quantitative satellite and radar data. *Preprints 7th Conf. Aerospace and Aeronautical Meteorology*, Melbourne, Amer. Meteor. Soc., 119–124.

Panofsky, H. A., and G. W. Brier, 1963: *Some Applications of Statistics to Meteorology*. The Pennsylvania State University Press, 224 pp. (see pp. 58–64).

Ray, P. S., 1976: Vorticity and divergence fields within tornadic storms from dual-Doppler observations. *J. Appl. Meteor.*, 15, 879–890.

Schlichting, H., 1968: *Boundary Layer Theory*. McGraw-Hill (see Chap. 24).

Woodward, B., 1959: The motion in and around isolated thermals. *Quart. J. Roy. Meteor. Soc.*, 85, 144–151.

Paper 38

Reprinted from JOURNAL OF APPLIED METEOROLOGY, Vol. 18, No. 4, April 1979
American Meteorological Society
Printed in U. S. A.

Thunderstorm Intensity as Determined from Satellite Data[1]

ROBERT F. ADLER

Laboratory for Atmospheric Sciences, NASA/Goddard Space Flight Center, Greenbelt, MD 20771

DOUGLAS D. FENN

GE/MATSCO, Beltsville, MD 20705

(Manuscript received 28 June 1978, in final form 15 December 1978)

ABSTRACT

Digital infrared data from a geosynchronous satellite (SMS 2) on 6 May 1975 are used to study thunderstorm vertical growth rates and cloud top structure in relation to the occurrence of severe weather (tornadoes, hail and high wind) on the ground. All thunderstorms from South Dakota to Texas along a north-south oriented cold front are monitored for a 4 h period with 5 min interval data.

An examination of five cloud elements having eight tornadoes indicates that in seven of eight cases the first report of the tornado took place during, or just after, a period of cloud top ascent. This vertical velocity is applicable to an area of 15 km on a side.

Thunderstorm growth rate, as determined by the rate of blackbody temperature isotherm expansion and minimum cloud top temperature, are shown to be correlated with reports of severe weather on the ground. A time analysis indicates that the derived parameters reach critical values soon enough to provide a potential warning lead time of approximately 30 min.

Equations are derived relating the thunderstorm growth rate to vertical velocity and outflow layer divergence. Severe thunderstorm elements are shown to have mean vertical velocities approximately twice as large as the non-severe elements. The outflow layer divergence is calculated to be 1×10^{-3} s^{-1} for the severe thunderstorms.

1. Introduction

Information obtained from geosynchronous satellite data has the potential of leading to greater understanding of convective and other mesoscale activity and also the potential of being an important component in the detection and monitoring of thunderstorm activity in general, and of severe thunderstorms in particular. The detection of thunderstorms using satellite data has centered on the interpretation of images. For example, Purdom (1976) has noted that intersecting cloud lines can produce enhanced convection and thunderstorms. However, quantitative information on cloud growth rates can also be determined by examining a sequence of images. Sikdar *et al.* (1970), Purdom (1971) and Arn (1975) have used ATS 3 visible data to measure thunderstorm anvil expansion rates. Adler and Fenn (1976), Yuen (1977) and Negri *et al.* (1976) have used SMS data to make similar calculations. Adler and Fenn (1976) showed examples of cold area expansion rates related to thunderstorm growth using SMS window channel infrared data. These measurements were for areas smaller than entire cirrus anvils and provide a technique to monitor individual thunderstorms rather than a group of storms under a large cirrus shield. This last study also included an example of a tornado preceded by the rapid expansion of a cold area, implying rapid ascent. All of these studies indicated that there are parameters determinable from geosynchronous satellite data that tend to be related to the occurrence of severe weather.

2. Objective and approach of study

The objective of the research reported on in this paper is to develop techniques to use geosynchronous satellite data to help detect and monitor severe thunderstorms, and to study their characteristics. Specifically, the purpose is to determine if thunderstorm growth rates and characteristics observable with SMS infrared (IR) data can be correlated with the occurrence of severe weather on the ground. The occurrence of thunderstorm-related severe weather (tornadoes, hail, high winds) is highly correlated with the intensity of convection. Using SMS IR data one should be able to calculate parameters related to convection intensity, such as cloud top temperature and its rate of change.

A particular area is outlined and all thunderstorms,

[1] A preliminary version of this paper was presented at the Tenth Conference on Severe Local Storms at Omaha, Nebraska.

0021-8952/79/040502-16$08.00

or thunderstorm clusters, in that area for a given time period are monitored. Parameters derived from the satellite data are then compared to ground reports of severe weather. The discussion of the results in this paper concentrates on examples of tornado-bearing thunderstorms and comparisons between severe and non-severe thunderstorms. Vertical velocity and divergence estimates are also presented.

3. Data and method of calculation

SMS 2 digital IR data from the period 1803–2208 GMT 6 May 1975 were used in this study. This period was part of a limited-scan observation period wherein a relatively narrow swath, with its long axis oriented east–west, was observed by the satellite, but on a frequent basis. For this case the nominal period between observations was 5 min. The spatial resolution of the IR channel is about 9.4 km at 40°N. There is a sampling overlap in the left-to-right direction so that each data point is centered 4.7 km from the center of the next. Lines are repeated during the data processing to complement the left-to-right oversampling so that image distances, either left-to-right or up-to-down, correspond approximately to each other in terms of geographical distance. Because of the sampling overlap in the left-to-right direction and the repetition of lines, each data point in this study represents an area approximately 4.7 km on a side, or 22 km^2.

The calculations presented in the next section were performed on the Atmospheric and Oceanic Information Processing System (AOIPS), an interactive image processing system described by Billingsley (1976). With this system, an irregularly shaped area can be outlined on a color enhanced television image and a histogram of the digital counts inside the area obtained.

The temperature resolution of the IR data is 1 K in the range of equivalent blackbody temperatures (T_{BB}'s) of interest in this study. Due to the fixed point algorithm used in the transformation of raw counts into calibrated values, voids appear in histograms (Goddard and Remondi, 1975). In the range of T_{BB} between 206–226 K, these voids appear every 3–5 K.

From the calculated histogram a cumulative histogram is compiled starting from the cold end of the temperature distribution. The number of points N_i with $T_{BB} \leq T_i$ is obtained, with T_i ranging from the lowest temperature in the area to the highest temperature T_0, where T_0 is defined as the warmest T_{BB} that forms a closed isotherm around the element being defined. In the case of easily definable, individual thunderstorms or thunderstorm clusters, an entire cirrus anvil may be defined. Examination of cirrus anvils led to the subjective determination that the $T_{BB} = 258$ K isotherm corresponds approximately to the anvil edge. Although this value is much warmer than the actual cloud temperature, the emissivity at the thin anvil edge is much less than unity, resulting in the warm T_{BB}. In addition, the response function of the sensor may result in a more gradual change in T_{BB} than is actually occurring. Preliminary examination of expanding anvils indicated that the areal change within somewhat lower T_{BB}'s paralleled the areal change within 258 K. For this reason a lower temperature, $T_{BB} = 226$ K, was chosen as the warmest cloud tops to be considered in the definition of areas. Therefore, the clouds studied here have already obtained a height of at least 9.5 km based on soundings for this day. The early development of the thunderstorms is thus not examined in this study. Satellite observations of earlier growth of thunderstorms is examined by Adler and Fenn (1978).

As the anvil of a thunderstorm or thunderstorm cluster expands, it eventually comes into contact with other anvils. When the warmer temperatures near the outer edge of the cloud can no longer be uniquely identified with the feature being examined, T_0 is decreased to a value such that the statistics of the data inside the area defined by T_0 now represent the feature being examined. In areas where a number of anvils have formed a continuous cirrus overcast, the object is to detect and monitor colder (higher) regions within the general cirrus background. In these cases T_0 is often very cold ($\sim$212 K) and the elements defined contain only a small range of T_{BB}.

The thunderstorms analyzed in this study lie along a generally north–south oriented cold front from Nebraska southward through Texas. This area contains all the reported tornadoes for this day during this time, other severe thunderstorms, and thunderstorms with no reported severe weather. A selection of SMS 2 images covering the time period studied is shown in Fig. 1. Especially late in the period there are areas of cirrostratus clouds which cannot be associated with individual thunderstorm anvils. Thunderstorm elements can still be identified by analyzing only the colder portions of a cloud. For example, Fig. 2 shows the distribution of selected T_{BB} isotherms for the region outlined in the 2128 GMT panel of Fig. 1. The area of interest in Fig. 2 is the region of relatively cold temperatures below the center of each panel. At 2057 GMT there is only a small area having a temperature of 208 K. Over the next 30 min the 208 K isotherm expands rapidly and colder temperatures appear. These changes imply ascent. The element defined by the 208 and 206 K isotherms is typical of the small, cold cloud elements monitored in this case. This particular element (defined as cloud 8b) was associated with the Omaha tornado which touched down at 2133 GMT. The position of the Omaha tornado relative to the cloud top T_{BB}'s is shown in the last panel of the figure. A hail report is shown in the panel for 2118 GMT. A discussion of tornado reports relative

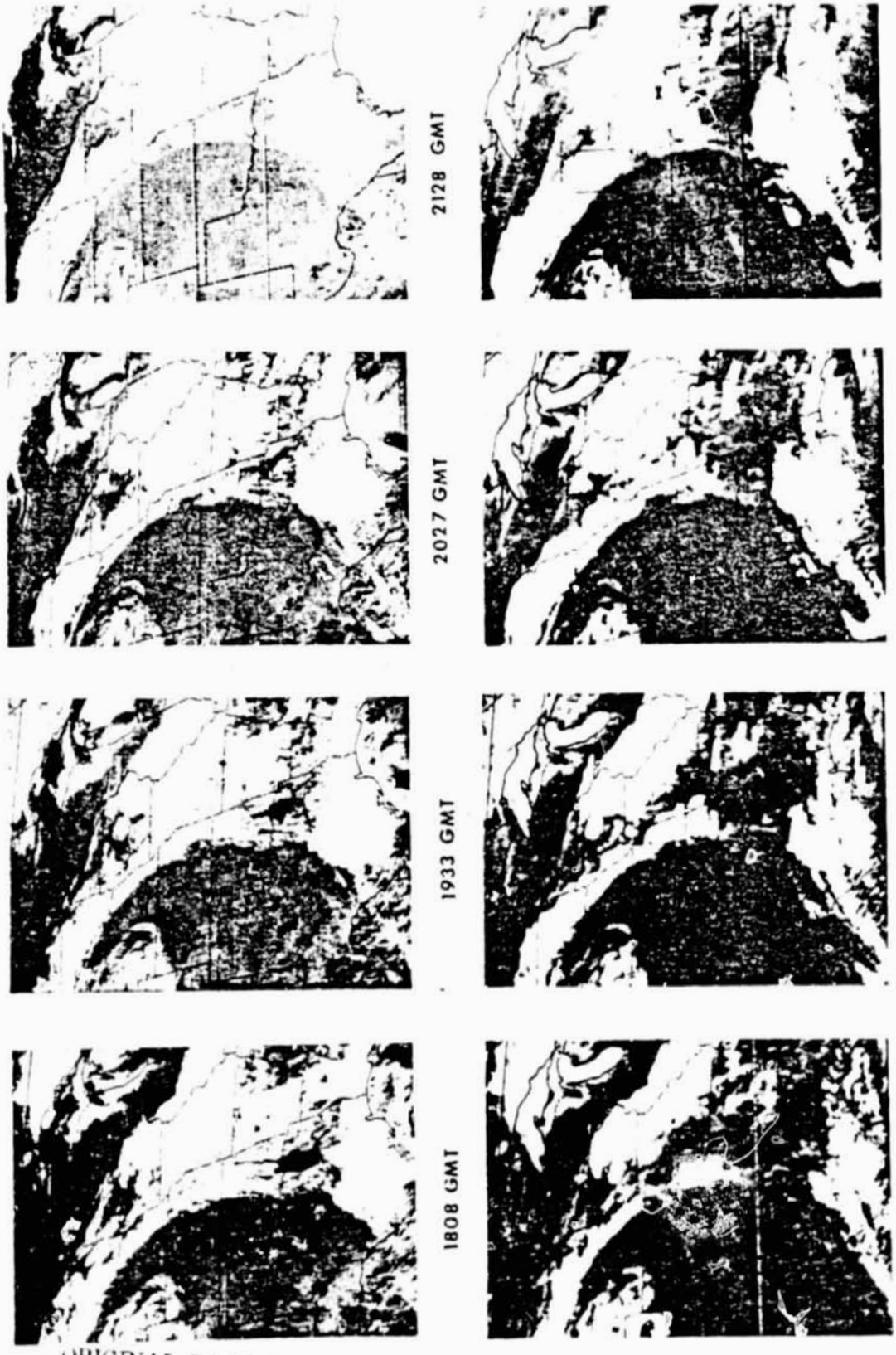


FIG. 1. Sequence of SMS 2 images over study area and period.

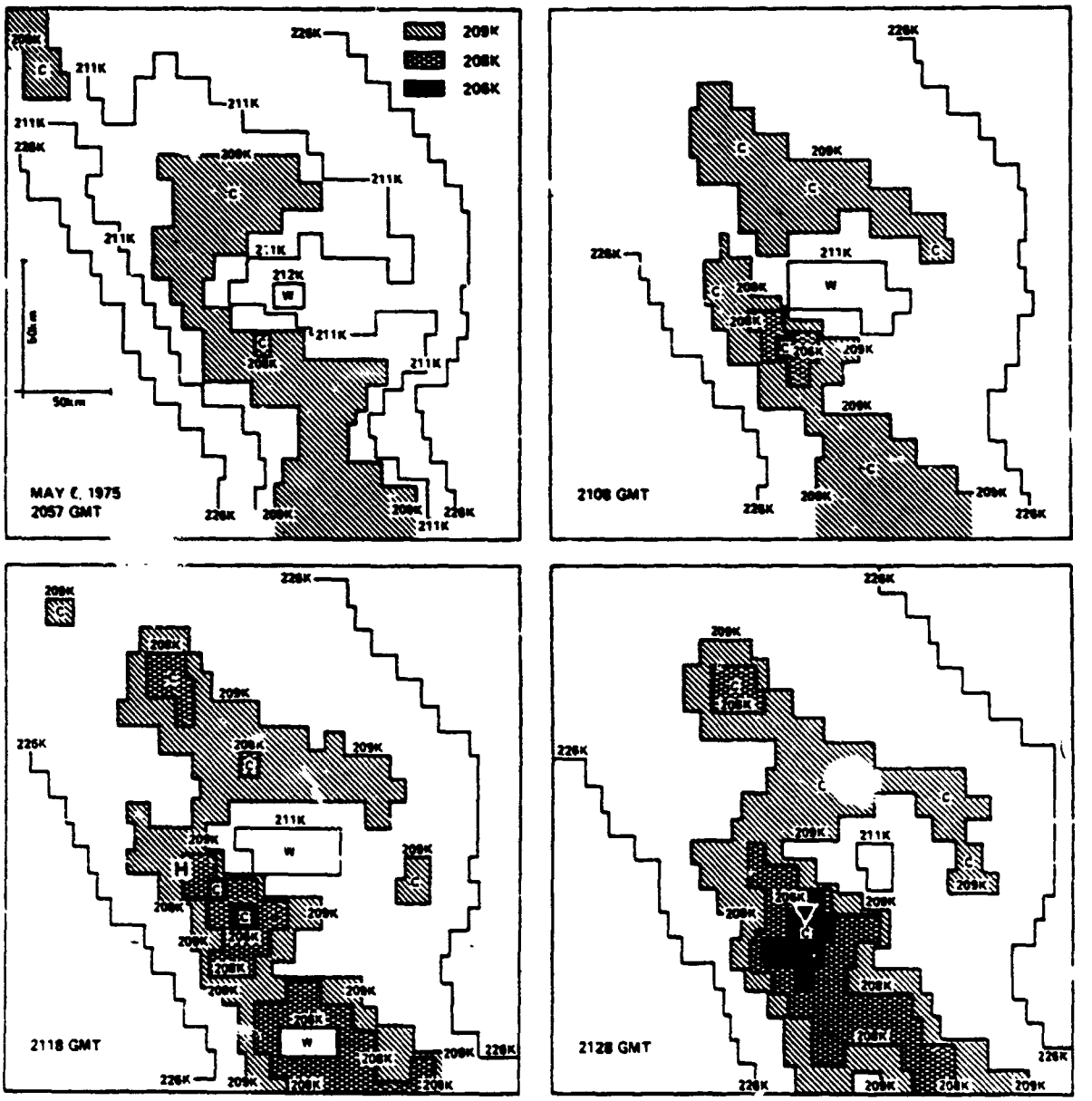


FIG. 2. Analyses of infrared T_{BB} isotherms over a small area shown in Fig. 1.
H marks location of hail reports; triangle marks tornado report.

to cloud top features is given in the next section. From Fig. 2 it can be seen that the expansion of the small cold areas cannot be interpreted as anvil expansion. Analysis of the cold area expansions in terms of vertical velocity and divergence is given in Section 7.

The occurrence of severe weather in this study is determined primarily from the Severe Weather Events Log compiled by the National Severe Storm Forecast Center (NSSFC) in Kansas City, with cross reference to the Environmental Data Service's *Storm Data*.

4. Results from tornado-bearing storms

Five clouds or elements in the area examined had tornadoes associated with them, with the total number of tornadoes being nine. The approximate locations of the tornadoes are given in Fig. 3. Four of these tornadoes can be considered relatively strong tornadoes. These are identified by nearby town names as the Pierce, Magnet, Winside and Omaha tornadoes, all occurring in Nebraska. The fifth tornado cloud is

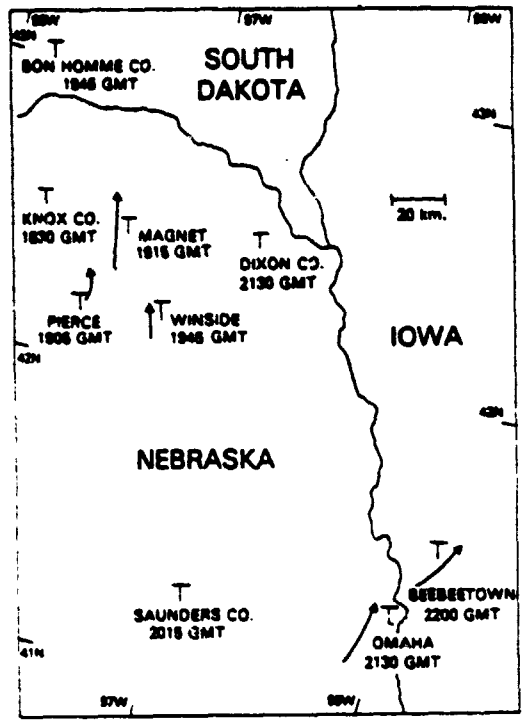


FIG. 3. Location of tornadoes on 6 May 1975 in area of interest. Arrows represent approximate tracks for relatively strong tornadoes.

associated with the relatively weak Saunders County (Nebraska) tornado.

Fig. 4 is the thunderstorm growth rate diagram for cloud 4. The cloud numbering system is the original one used in the study to identify approximately 60 clouds or elements from South Dakota to Texas. Eventually, only 39 elements were determined to be active thunderstorm areas. The eliminated elements could only be followed for a short time and indicated little or no growth. The cloud or element numbers will, therefore, not necessarily be sequential. The diagram indicates the relation between N, and time, where N, is the number of data points in the defined element with blackbody temperature $T_{BB} \leq T$,. For example, the curve in Fig. 4 for T, = 218 K exhibits an increase from N = 14 at 1808 GMT to N = 60 at 1823 GMT. Before 1808 the minimum T_{BB} was warmer than 218 K. The curve is drawn to N = 1 at 1803 GMT, so that the growth from zero to 14 can be displayed. Because the N axis is logarithmic, the slopes of the lines are proportional to $N^{-1}dN/dt$. Plotted points which are circled indicate that the area is not defined uniquely by that isotherm, but that a portion of the border is approximated. The approximation is usually small and does not significantly affect the count. The conversion from number to area is accomplished by multiplying the number of points by 22 km^2. The location in time of the severe weather events is shown across the top of the diagram.

Cloud 4 in Fig. 4 shows a rapid decrease in minimum blackbody temperature, or T_{min}, and a rapid expansion of areas within isotherms. The decrease in temperature can be calculated by moving horizontally across the diagram at, for example, N = 10. The 216 and 212 K lines are separated by 20 min, indicating a rate of decrease of 0.2 K min^{-1}. The values of $N^{-1}dN/dt$ are of order of magnitude 10^{-3} s^{-1}. The decrease in temperature and expansion of isotherms, of course, implies ascent. The vertical velocity is related to the rate of temperature change by

$$w = \left(\frac{\partial T}{\partial z}\right)^{-1} \frac{dT}{dt}. \qquad (1)$$

With a reasonable lapse rate, 8 K km^{-1}, the temperature change of 0.2 K min^{-1} is equivalent to a vertical velocity of w of 0.4 m s^{-1}. This is a reasonable vertical velocity estimate considering the area over which it is an average (~225 km^2) and considering it represents a time when the cloud top is decelerating as it approaches and penetrates the tropopause at approximately 12 km. Much larger values of vertical velocity are calculated for clouds whose tops are just 1 or 2 km lower. A more detailed examination of the vertical velocity estimates is given in Section 7.

Three tornadoes are associated with cloud 4 (Fig. 4). One in Bon Homme County (South Dakota) is in the

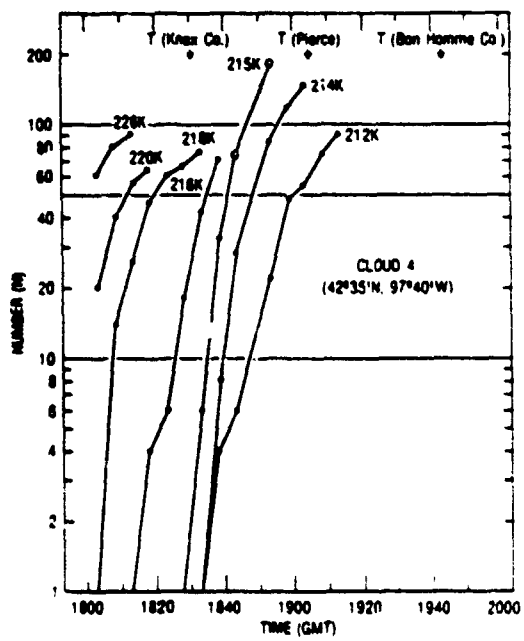


FIG. 4. Thunderstorm growth rate diagram for cloud 4. N is the number of data points in the element with blackbody temperature $T_{BB} \leq T$,, where T, is as identified in the diagram. Occurrence of severe weather is marked by arrow at time of report, with T indicating tornado.

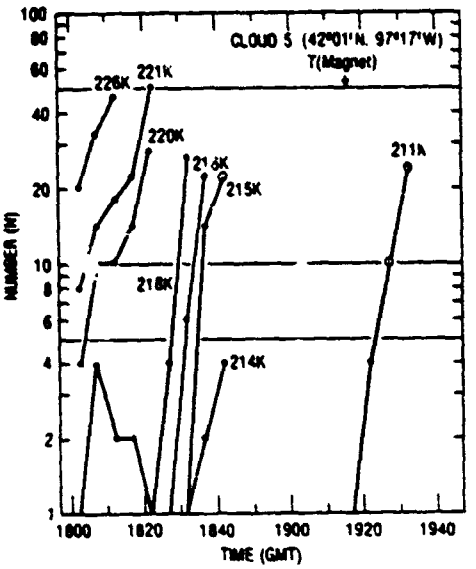


FIG. 5. Thunderstorm growth rate diagram for cloud 5.

vicinity of this element, but when no distinct cold area can be defined. The Pierce tornado, the strongest of the three, was first reported at 1905 GMT. At that time and immediately preceding it, during the tornado formation period, the element exhibits cold area expansion and decreasing temperature, indicating ascent. The same pattern also is associated with the Knox County tornado.

Fig. 5 shows the diagram for cloud 5 which is identified as having been associated with the Magnet tornado. At the beginning of the period, relatively slow growth is indicated for the 221 and 220 K areas. A few points with $T=218$ K appear and then disappear before a period of very rapid growth is observed between 1825–1840 GMT. The maximum of $N^{-1}dN/dt$ is 6.2×10^{-3} s^{-1} at 218 K. The minimum temperature also drops very rapidly during this period. After 1843 GMT no clearly defined cold area can be identified, although the cloud element can still be identified in the higher resolution visible images. In the IR data cloud 5 appears as an extension of an element immediately to the south, cloud 22. Not until 1923 GMT or 40 min later, is a definable cold area at 211 K observed for cloud 5. Over the next 10 min this isotherm expands swiftly. The initial report of the tornado is at 1915 GMT, as noted on the diagram, but the tornado hits the town of Magnet at 1945 GMT and lasts at least until 2000 GMT. It, therefore, appears that this tornado also formed and touched down during a time when the thunderstorm top was ascending at the spatial scale observable with the IR data.

Just to the south of the Magnet cloud is cloud 22 (Fig. 6). This element can be defined continuously for almost 3 h. Two tornadoes are associated with this element at two widely separated times. The first, at Winside, Nebraska, is reported at 1945 GMT. As with the tornadoes already examined, this tornado occurs during a period of rapid increase in cold area, which in this case lasts from about 1900 GMT until 2000 GMT. The maximum value of $N^{-1}dN\ dt$ is 4.5×10^{-3} s^{-1}, following the 211 K isotherm. The second, weaker

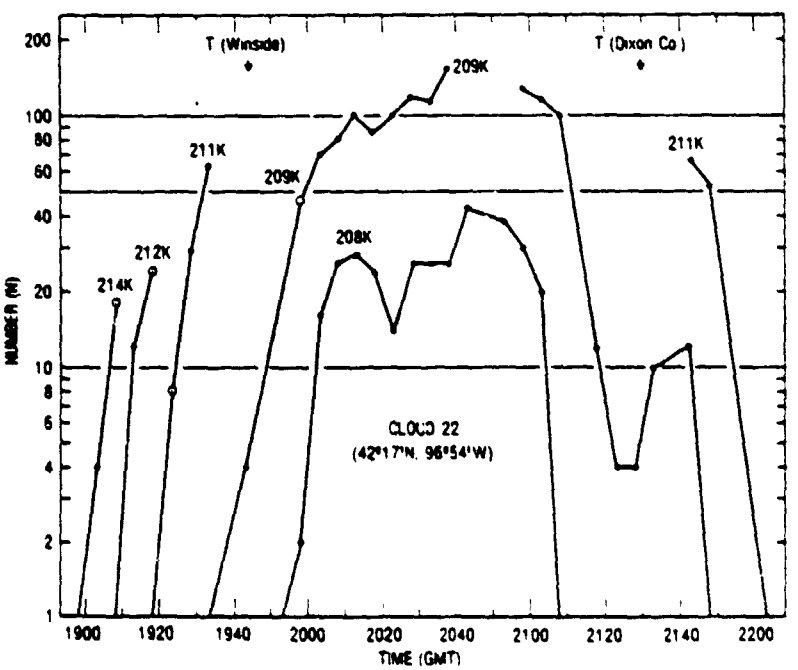


FIG. 6. Thunderstorm growth rate diagram for cloud 22.

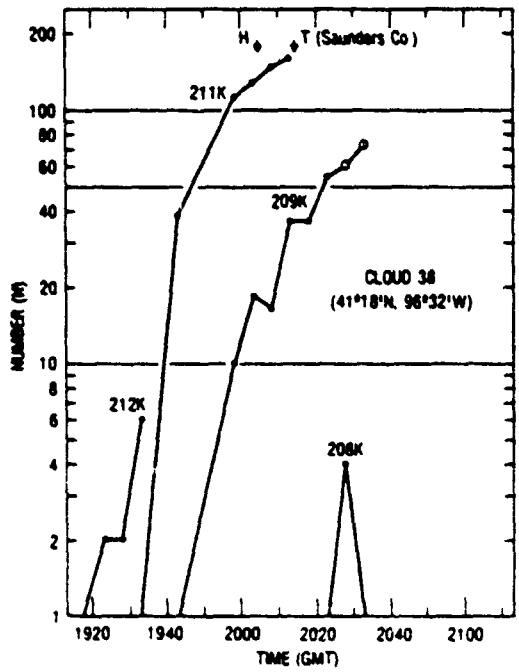


FIG. 7. Thunderstorm growth rate diagram for cloud 38. H indicates report of hail.

tornado, at 2130 GMT, is located in Dixon County, Nebraska. Although there is a brief period of expanding cold area between 2125–2145 GMT, this tornado occurs during a period of generally decreasing cold areas and warming temperatures. One can speculate that during the dying phase of element 22, one last vigorous thunderstorm penetrates the cirrus shield, showing up for a short time as the secondary maximum in the 209 K curve at 2140 GMT.

A tornado occurring in Saunders Country, Nebraska, is associated with cloud 38 (Fig. 7). Again, the reported time of the tornado is during a period of cold area expansion indicating ascent. The most rapid increase in N is between 1930–1950 GMT at a rate of 6.7 $\times 10^{-4}$ s^{-1}. The minimum temperature reached is 208 K.

The curves for cloud 8b are shown in Fig. 8. Two tornadoes are associated with this cloud, including the severe Omaha tornado. This element is observed to go through two periods of rapid ascent during a 2 h period. Reports of hail and high wind (>50 kt) are associated with the first period of rapid expansion. The element is then observed to have a decrease in cold area between 2040–2055 GMT. After that there is a sharp increase in N(208 K) and colder temperatures appear. Although there is good continuity of the element during this cycle of inferred ascent and descent, the second period of ascent is probably brought about by a new thunderstorm penetrating the cirrus shield in approximately the same area.

Similar to previous cases, the growth rate of cloud 8b during the 30 min period prior to the first tornado report indicates ascent. During the lifetime of the tornado (2133–2150 GMT) no very sharp area increase is noted. In fact, the dip in the 206 K line could possibly be interpreted as a partial collapse of the thunderstorm top. This is the only element for which such a feature is evident. No data are available from 2150 to 2208 GMT. At the later time, 205 K temperatures appear, indicating more growth. The Beebeetown, Iowa, tornado is reported during the period 2200 to 2230 GMT.

During the time period examined, nine tornadoes were reported, eight of which could be clearly connected with a small, cold area in the IR data. The exception was the Bon Homme County tornado which was located in the vicinity of cloud 4, but long after it was no longer possible to define a small, distinct cold area in it. Of the remaining eight cases, all but one (Dixon County—cloud 22) occurred during, or just after, a rapid increase in cold area, indicating cloud top ascent. Therefore, it appears that the formation of tornadoes in a large majority of cases occurred during a period of increasing thunderstorm cloud top height. This upward vertical velocity applies on a spatial scale of approximately 15 km on a side.

The observations of increasing thunderstorm top heights at the time of, or just prior to, tornado touchdown in the present study seemingly contradict aircraft observations of overshooting tops discussed by Fujita (1973), Pearl (1974) and Umenhofer (1975). They relate collapsing cloud domes (above the anvil cirrus) to the occurrence of tornadoes. This discrep-

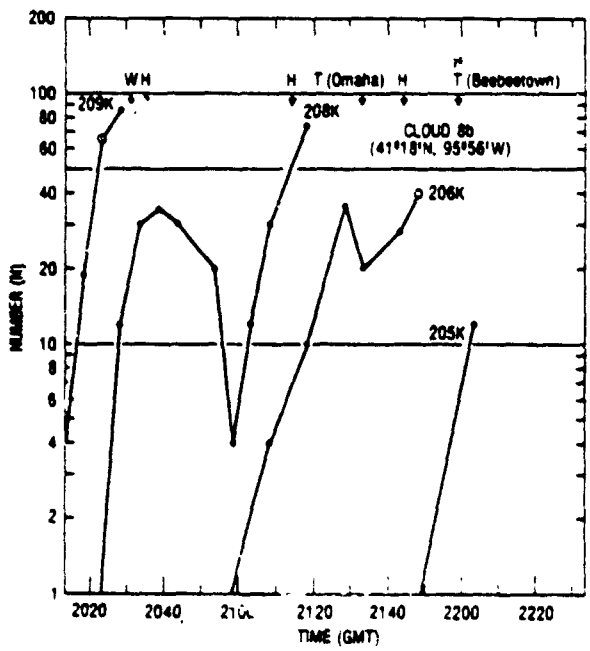


FIG. 8. Thunderstorm growth rate diagram for cloud 8b. W indicates high winds.

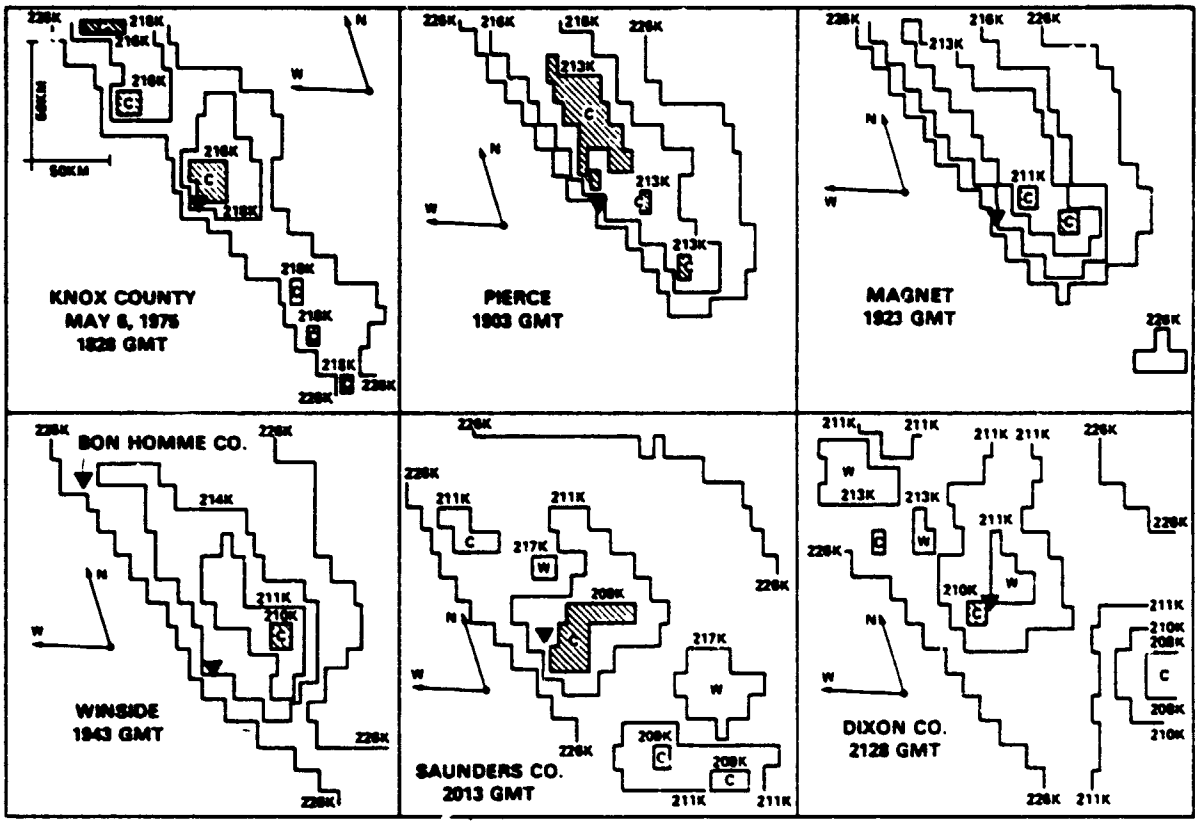


FIG. 9. Location of tornado reports relative to IR cloud top features.

ancy may be related to a difference in scale between the two types of observations. The satellite observations apply to a scale of about 15 km on a side. Shenk (1974) has examined the size distribution of overshooting domes or turrets and finds a median diameter of approximately 5 km. Therefore, the satellite vertical motions are applicable to an area approximately an order of magnitude larger than that observed with the aircraft. Comparison of results is therefore difficult.

Radar observations also do not show a sharp increase in echo top height prior to tornado touchdown. For example, Lemon *et al.* (1978) show nearly steady or slightly decreasing radar heigh s for the 30 min prior to the Union City, Oklahoma, tornado. A rapid decrease of the radar top follows the tornado touchdown. Again these observations are on a small scale compared to the satellite-based cloud top changes. A full understanding of the relation of satellite-inferred cloud top changes, radar echo height and actual cloud changes on various scales must await closer coordination in the observations.

Locations of six tornado reports relative to cloud top features observed in the satellite IR data are shown in Fig. 9. A similar diagram for the Omaha tornado is given in Fig. 2. The surface tornado locations have been shifted to cloud top relative positions for a cloud height of 12 km to compensate for the satellite viewing angle. The tornado reports tend to be located to the west or southwest of the coldest temperatures, in an area of large gradient of T_{BB}. Early in the 4 h period (before 2000 GMT) the tornado reports are located near the 226 K T_{BB} isotherm, which is near, but inside, the upwind cirrus anvil edge. Later, after the cloud has expanded and the anvil edge is farther from the cold center, the tornado reports are located more distant from the cloud edge.

5. Example comparison of a severe and non-severe case

Two adjacent thunderstorm elements which developed in southeast Nebraska at about the same time provide an excellent comparison of clouds with and without accompanying severe weather. The diagrams for these elements (clouds 8 and 9) are shown in Figs. 10 and 11, respectively.

Both elements reach a temperature of 226 K a little after 1800 GMT and continue to grow to higher heights and colder temperatures during the next two hours. However, there are significant differences in the rate of growth and the final temperature reached. Cloud 9, with no accompanying severe weather reports, has lower values of normalized growth rate $N^{-1}dN/dt$ and an associated less rapid decrease of minimum

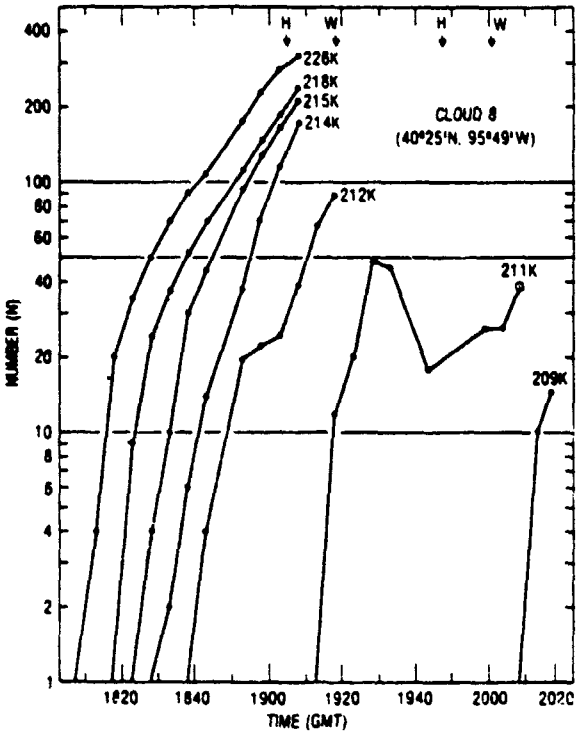


FIG. 10. Thunderstorm growth rate diagram for cloud 8.

temperature. Its eventual coldest temperature is 212 K compared to 209 K for cloud 8. Two reports of hail and two reports of high wind are associated with cloud 8. The first pair of hail and wind reports occur near the end of the period of rapid growth; the second pair happen during a relatively stable period. This second pair may be due to a thunderstorm growing rapidly from below, but hidden by the already high and dense cirrus anvil of the previous cell. Finally, at the end of the period the new thunderstorm penetrates the anvil and becomes observable as the 209 K area at 2015 GMT.

The comparison of cloud elements 8 and 9 is the best available of those analyzed. These two clouds began about the same time, lasted a substantial amount of time and were easily defined. Other elements were sometimes obscured by anvils of more rapidly growing storms and other weak elements often became undefined in a general cirrostratus shield. In general, however, the strong thunderstorms are easily observable.

6. Results for all storms

Results from the analysis of 39 thunderstorms or thunderstorm clusters are discussed in this section. There are 15 elements associated with severe weather and 24 with no reported severe weather. In the area analyzed these 39 elements are all the clouds defined as growing thunderstorms. A few other elements were defined but lasted only a short time, showed minimal change, and no severe weather reports were associated with them.

An examination of the thunderstorm growth diagrams and the patterns of digital data gave the impression that two simple characteristics were correlated with the occurrence of severe weather. These characteristics are minimum cloud top temperature and thunderstorm growth rate as given by the normalized rate of cold area expansion.

Fig. 12 displays the results of plotting the minimum temperature during an element's lifetime (T_{min}) against the thunderstorm growth rate. The dots represent non-severe cases, while T's and H's represent tornadoes and hail respectively. The following procedure was used to arrive at a single value of $N^{-1}dN/dt$ to represent each element. The maximum $N^{-1}dN/dt$ value, for any temperature $\leqslant 226$ K, is determined by examining the time change between $N=6$ and $N=30$ for each temperature. These two values were chosen so that the rate of increase is for relatively small areas, but that there are enough data points (6) to establish the lower threshold without being affected severely by instrument noise. A few elements had no curves of N lasting from $N=6$ to 30, so the parameter was calculated using the largest range available.

An examination of Fig. 12 indicates that the severe weather elements tend to have cold minimum cloud top temperatures and large rates of growth. A series of discriminant analyses (Panofsky and Brier, 1963) were performed, using the data displayed in Fig. 12. Table 1 shows the results. For T_{min} by itself a severe,

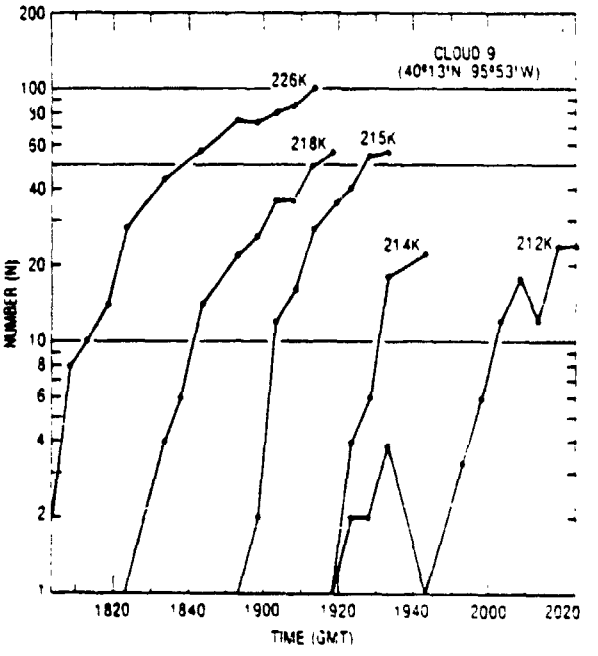


FIG. 11. Thunderstorm growth rate diagram for cloud 9

non-severe dividing line of 212.8 K was derived. For temperatures above 213 K only one out of 12 elements had severe weather, while 14 out of 27 with temperatures colder than 213 K had associated severe weather reports. For maximum rate of growth the derived dividing line is 3.5×10^{-3} s^{-1} with severe and non-severe events distributed around that value as shown in the table. A combined discriminant analysis using T_{min} and $N^{-1}dN/dt$ produced the results shown in the last row in the table and by the diagonal line in Fig. 12. Based on the two variable discriminant analysis a probability of detection (POD) equal to 0.73 was calculated along with a value of 0.31 for the false alarm rate (FAR) and 0.55 for the critical success index (CSI). The definition of these variables follows Donaldson *et al.* (1975). There are five elements with associated tornadoes, and their relations to the derived critical values are given by the fractions in parentheses in Table 1.

Statistical tests were also performed to determine if there was a significant difference between the means for severe thunderstorms and non-severe thunderstorms. For T_{min} the mean values were 209.5 and 212.3 K for severe and non-severe cases. Using a t-distribution test, the difference of means was shown to be significant at the 5% level, with the value of t just missing the value for the 1% level. For $N^{-1}dN/dt$ the means were 4.5×10^{-3} and 3.1×10^{-3} s^{-1}. This difference of means also is significant at the 5% level, again barely missing the 1% value.

It is not surprising that T_{min} and thunderstorm growth rate ($N^{-1}dN/dt$) are correlated with occurrences of severe weather. Both parameters are obvious indicators of the intensity of convection. Colder temperatures imply higher thunderstorm heights, which have long been related to thunderstorm severity through radar observations. All 39 elements involved in this analysis were associated with vigorous thunderstorms. The very weakest, as indicated in the bottom, left-hand corner of Fig. 12, had a minimum temperature of 219 K, which is equivalent to a height of 11.1 km, using the nearby 0000 GMT 7 May Omaha sounding. The coldest cloud had a temperature of 205 K, or a height of 13.0 km. The 0000 GMT 7 May

TABLE 1. Discriminant analysis summary for cloud top minimum temperature, maximum rate of growth, and the two variables combined. The parameter L is positive for predicted severe weather.

Variable(s)	Value for $L=0$	Fraction of elements with severe weather* $L<0$	$L>0$
Cloud top minimum temperature (T_{min})	212.8 K	$\frac{1}{12}\left(\frac{0}{12}\right)$	$\frac{14}{27}\left(\frac{5}{27}\right)$
Maximum rate of growth $\left(\frac{1}{N}\frac{dN}{dt}\right)$	3.5×10^{-3} s^{-1}	$\frac{5}{22}\left(\frac{1}{22}\right)$	$\frac{10}{17}\left(\frac{4}{17}\right)$
Combined T_{min} and $\frac{1}{N}\frac{dN}{dt}$	$L=9.75+91\left(\frac{1}{N}\frac{dN}{dt}\right)-0.048\ T_{min}$	$\frac{4}{23}\left(\frac{0}{23}\right)$	$\frac{11}{16}\left(\frac{5}{16}\right)$

* Values in parentheses indicate tornado cases.

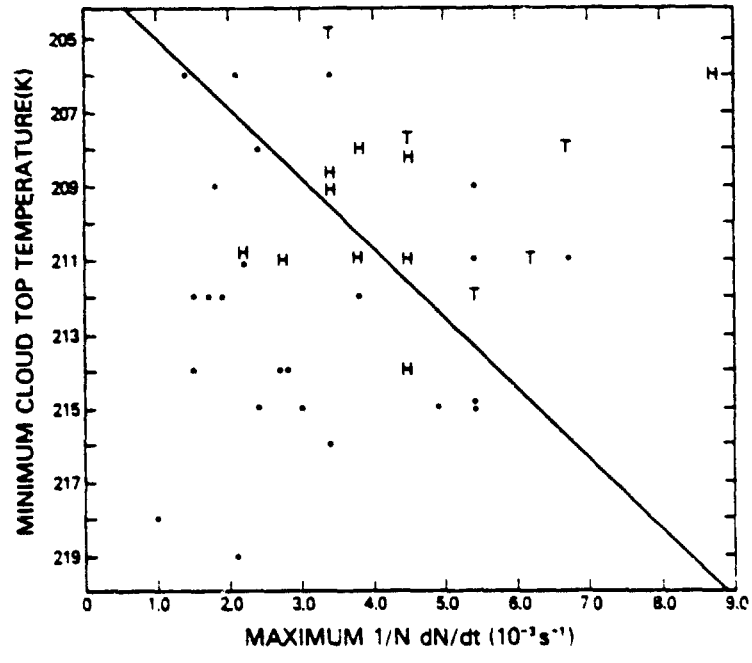


FIG. 12. Relation of thunderstorm rate of growth (cold area expansion) to cloud top minimum temperature and the occurrence of severe weather for all 39 thunderstorm elements in analysis area. T indicates tornado H represents reported hail.

TABLE 2. Median warning lead time for all elements with severe weather.

Criterion	Number of cases	Median value of lead time (min)
T_{min}	14	24
$\frac{1}{N}\frac{dN}{dt}$	10	27.5
Combined	11	7

sounding has a minimum temperature of 206 K. The tropopause is probably located at 222 mb at a temperature of 215 K, although there is another sharp increase in static stability above that, at a pressure of 168 mb and at a minimum temperature of 206 K. Twelve hours earlier the Omaha sounding has the tropopause located at 200 mb and 213 K. Farther south along the area studied the tropopause temperature was also in the vicinity of 213 K. At Monett, Missouri, the tropopause temperature values were 211 and 215 K at 1800 GMT 6 May and 0000 GMT 7 May. Thus an approximate tropopause temperature for this region on this day was 213 K, which is the threshold temperature in the discriminant analysis discussed earlier. This result may be accidental, but is supported by similar observations by Pryor (1978). Additional studies are needed to ascertain the relations among the temperature structure and tropopause height, the satellite-observed cloud-top temperatures and storm severity.

The thunderstorm heights estimated from the satellite data in this study are substantially lower than the corresponding radar estimates. A cursory comparison of the satellite-based heights with the radar-estimated storm tops as analyzed on the National Weather Service's radar summary chart, indicates a mean 2 km underestimate from the satellite information.

The reason for the radar-satellite differences is probably related to the scale of the observations and the limitations of the satellite data. The satellite-observed cloud top temperatures are for a larger horizontal area. The instantaneous field of view (IFOV) of the satellite IR sensor at this latitude is approximately 9 km on a side. The radar observations are probably representative of a smaller region.

The satellite data are also limited by the temperature structure. For elements of the size with which we are dealing, the cloud top probably takes on a temperature which is a combination of updraft temperature (from the moist adiabat) and the ambient air at that altitude. This will have the effect of somewhat limiting the greatest indicated height. Clouds penetrating into the lower stratosphere will have their height underestimated.

In addition, Negri *et al.* (1976) have already noted that SMS IR temperatures underestimate thunderstorm height, especially for small elements. They attribute this effect to inadequate sensor response when going from a warm to a cold target. This is also a probable contributor to the radar-satellite difference noted in this study. Because of these factors the satellite-based estimates of thunderstorm height should be treated cautiously, although relative heights are probably valid.

The thunderstorm growth rate parameter $N^{-1}dN/dt$ is an indicator of the magnitude of the upward vertical velocity and the upper level divergence. The relation of these three variables is explored in the next section. One interesting point from Fig. 12 is that T_{min} and $N^{-1}dN/dt$ appear to be almost uncorrelated. One would expect that faster growing storms would penetrate to higher heights and colder temperatures, and therefore that the two parameters would be correlated. It is believed that this lack of correlation arises because the thunderstorm elements in many cases were hidden by thick cirrus produced by other convection and, consequently, could not be monitored early in their history when the growth rates were most likely the most rapid. By the time the element penetrates the cirrus shield and growth rates can be calculated, the rate of growth has slowed. Such elements reaching cold temperatures, however, are represented by points on Fig. 12 with low temperatures and low growth rates. Another problem is the comparison of $N^{-1}dN/dt$ values irrespective of at what temperature or height they are calculated. Thus values of $N^{-1}dN/dt$ of 5×10^{-3} s^{-1} calculated at $T_{BB}=220$ K may not reflect as large a relative velocity as 5×10^{-3} s^{-1} at 210 K. With the addition of more cases these types of refined analysis will be possible.

The parameters calculated appear to have the potential to be positive contributors to a severe storm warning system. To test what type of warning lead times might be achieved, the following calculations were made. Based on the criteria in Table 1, the time difference between when the element met the criterion and the time of the first report of severe weather was calculated for each element that met the criterion. The median values are displayed in Table 2. The median lead times are ~25 min for the one-variable criteria, but only 7 min for the combined criterion.

TABLE 3. Median warning lead time for elements with tornadoes.

Criterion	Number of cases	Median value of lead time (min)
T_{min}	5	33
$\frac{1}{N}\frac{dN}{dt}$	4	35
Combined	5	29

A second calculation was made using the first report of tornadoes only and the results are shown in Table 3. The median lead time is ~30 min. The results of Tables 2 and 3 point out that the parameters calculable from the SMS data are not only correlated with reports of severe weather, but usually meet the severe thunderstorm criteria before the actual severe weather (hail, tornadoes) occurs.

7. Vertical velocity and divergence estimates

a. Interpretation of $N^{-1}dN/dt$

The parameter chosen in this study to indicate thunderstorm rate of growth is the areal expansion of blackbody temperature isotherms in the window channel infrared data. This parameter has been chosen in place of rate of temperature change because many instances arise where this rate is difficult to calculate due to poor temperature resolution. Figs. 7 and 8 are examples of this situation. The calculated values of $N^{-1}dN/dt$ also appeared to be relatively independent of the temperature or height.

If a blackbody temperature isotherm is chosen so that it nearly coincides with the edge of the thunderstorm anvil, normalized expansion of the area within the isotherm is a direct measure of outflow divergence. That is,

$$D\ (\text{divergence}) = \frac{1}{A}\frac{dA}{dt}, \tag{2}$$

disregarding any dissipation of the anvil edge. In the present study, however, the isotherm expansions ($N^{-1}dN/dt$) calculated were usually chosen for areas well within the anvil edge. There are two reasons for this choice. First, the emphasis here is on obtaining parameters which are applicable on a spatial scale as close as possible to thunderstorm scale. This is why the $N^{-1}dN/dt$ values in Fig. 12 are calculated between $N=6$ and $N=30$. The anvil would cover an area two orders of magnitude larger. Second, it is often difficult to define an anvil edge because of complex interactions between thunderstorms.

The calculations of $N^{-1}dN/dt$ from $N=6$ to $N=30$ are applicable to an area of about 15 km on a side and definitely do *not* represent anvil expansion. The results of this section will show that vertical velocity estimates can be made from the isotherm expansion values.

The interpretation of $N^{-1}dN/dt$ in terms of vertical velocity and divergence is shown schematically in Fig. 13. The cross-sectional or side view in upper part of the diagram shows the curved surface of the cloud top at two times, t_1 and t_2. The vertical coordinate is temperature. The emissivity of the cloud top is assumed to be unity. This should be a good assumption in the inner part of the thunderstorm top where this analysis applies. From times t_1 to t_2 the

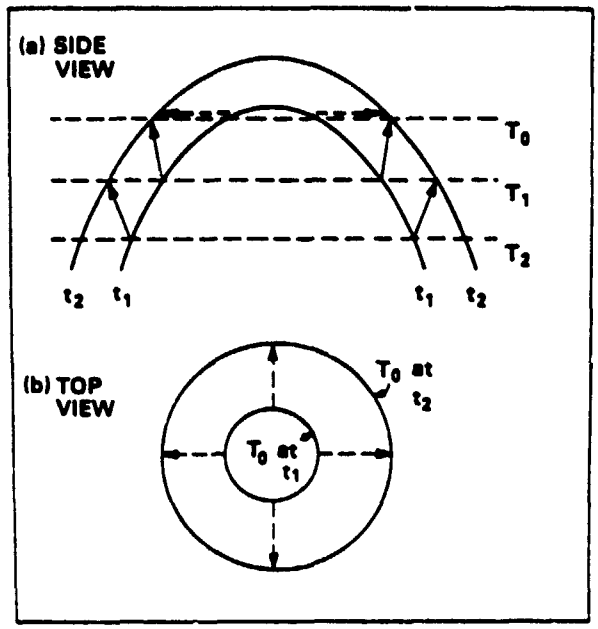


FIG. 13. Schematic views of idealized cloud top at two different times. Temperatures are indicated by T_0, T_1 and T_2. Solid arrows represent actual air motion, while dashed vectors represent the expansion of blackbody temperature isotherms, as viewed by the satellite.

area inside the T_0 isotherm expands. This is seen by the horizontal dashed arrows in the upper part of the figure and is also indicated in the top view in the lower part of the diagram. This areal expansion is equivalent to the $N^{-1}dN/dt$ which is measured with the satellite data.

The actual flow of air, however, is given by the solid arrows The air ascends and therefore cools (moist adiabatically) and is somewhat divergent. If this simple schematic diagram is basically correct then the expansion of cold areas as denoted by $N^{-1}dN/dt$ is really a combination of vertical velocity and divergence effects.

b. Derivation of equations

The relationships of $N^{-1}dN/dt$, divergence and vertical motion can be shown through the following derivation. A point on the cloud surface is represented by the position vector

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}, \tag{3}$$

and the motion of the point is given by

$$\frac{d\mathbf{r}}{dt} = \mathbf{V} = \frac{dx}{dt}\mathbf{i} + \frac{dy}{dt}\mathbf{j} + \frac{dz}{dt}\mathbf{k}. \tag{4}$$

The horizontal motion in the x direction (Fig. 13a is in the x-z plane) of a point on the cloud surface is

given by

$$u=\frac{dx}{dt}=\frac{\partial x}{\partial t}\bigg|_z+w\frac{\partial x}{\partial z}\bigg|_t, \tag{5}$$

where $\partial x/\partial t|_z$ is the horizontal movement of the cloud edge at a particular altitude or temperature. In Fig. 13a this is represented by the horizontal, dashed vector. The term $\partial x/\partial z|_t$ is the slope of the cloud surface. Differentiating Eq. (5) with respect to x and adding the result to a similar expression for v differentiated with respect to y results in

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=\frac{\partial}{\partial x}\left(\frac{\partial x}{\partial t}\bigg|_z\right)+\frac{\partial}{\partial y}\left(\frac{\partial y}{\partial t}\bigg|_z\right)+w\nabla_z\cdot\frac{\partial r}{\partial z}, \tag{6}$$

if w is assumed to be constant in the horizontal. The left-hand side of Eq. (6) is the horizontal divergence D and the sum of the first and second terms on the right side is equivalent to the rate of isotherm expansion and the sum is equal to $N^{-1}dN/dt$. Therefore Eq. (6) can be expressed as

$$D=\frac{1}{N}\frac{dN}{dt}+w\nabla_z\cdot\frac{\partial r}{\partial z}. \tag{7}$$

The second term on the right side of Eq. (7) is a product of the vertical velocity and a divergence term which is a function of the slope of the cloud surface. By incorporating a vertical lapse rate into the expression, the vertical coordinate is shifted to temperature in Eq. (8):

$$D=\frac{1}{N}\frac{dN}{dt}-w\left(-\frac{\partial T}{\partial z}\right)\nabla_z\cdot\frac{\partial r}{\partial T}. \tag{8}$$

Information about the shape of the cloud top can be inferred from the satellite data, by determining the area encompassed by the then coldest isotherm when the next, still colder isotherm appears. If the area of the first isotherm is large the slope is small and if the area is small, this implies a relatively steep slope.

By defining a shape parameter s, such that

$$s=\left[\nabla_z\cdot\frac{\partial r}{\partial T}\right]^{-1}, \tag{9}$$

one modifies Eq. (8) to become

$$D=\frac{1}{N}\frac{dN}{dt}-\frac{w}{s}\left(-\frac{\partial T}{\partial z}\right). \tag{10}$$

Using Eq. (1) to substitute into Eq. (14) a useful alternate expression,

$$D=\frac{1}{N}\frac{dN}{dt}+\frac{1}{s}\frac{dT}{dt}, \tag{11}$$

is derived. Eq. (10) states the relation among divergence, isotherm expansion, and vertical velocity. As will be shown, divergence is small compared to the other terms. The parameter s is large when the cloud has steep slopes and small when the slope is shallow. An expression for s, which can be used with SMS/GOES IR data, is derived in the next section. Although not used in this study, information from the higher resolution visible channel could be used to help infer cloud top shape during daylight hours, especially near sunrise and sunset.

c. Formulation for parameter s

Information concerning the shape of the cloud top and therefore the shape parameter s can be determined from the satellite IR data. The term in Eq. (9) is first expanded to obtain

$$\nabla_z\cdot\frac{\partial r}{\partial T}=\frac{\partial}{\partial x}\left(\frac{\partial x}{\partial T}\bigg|_t\right)+\frac{\partial}{\partial y}\left(\frac{\partial y}{\partial T}\bigg|_t\right). \tag{12}$$

Assuming symmetry in the two directions results in

$$\nabla_z\cdot\frac{\partial r}{\partial T}=2\frac{\partial}{\partial x}\left(\frac{\partial x}{\partial T}\bigg|_t\right). \tag{13}$$

The slope $\partial x/\partial T$ is approximated by

$$\frac{\partial x}{\partial T}\approx\frac{\xi N_w^{-1}}{2\Delta T}, \tag{14}$$

where ξ is the length represented by each satellite data point or pixel and N_w the number of data points at T_1 when the next coldest temperature T_2 first appears. The expression assumes a square area. Therefore $\xi N_w^{-1}/2$ is an estimate of the distance from the center of the isotherm area to the edge of the area. Divided by ΔT this produces an estimate of the slope in Eq. (14). The term ΔT is not always 1 K but is often approximated by 1.2 or 1.3 K because of the temperature void problem mentioned in Section 3.

The finite difference approximation for Eq. (13) is simply

$$2\frac{\partial}{\partial x}\left(\frac{\partial x}{\partial T}\right)\approx 2\frac{(\partial x/\partial T)_1-(\partial x/\partial T)_2}{\Delta x}, \tag{15}$$

where the derivatives at positions 1 and 2 are the slopes at opposite sides of the cloud top, and Δx is the width of the area enclosed by the isotherm. Assuming they are equal in magnitude and opposite in sign and substituting Eq. (14) for the slope gives

$$2\frac{\partial}{\partial x}\left(\frac{\partial x}{\partial T}\bigg|_t\right)=\frac{2\xi N_w^{-1}}{\Delta T\Delta x}. \tag{16}$$

Because the calculations to follow are for areas of

the size of $N=6$ to $N=30$ satellite data points, an intermediate value of $N=16$ was chosen as an appropriate size for the application of these formulas. Therefore, Δx in Eq. (16) is assigned a value of $\sqrt{16}\xi$ or 4ξ. Thus

$$\nabla_z \cdot \frac{\partial \tau}{\partial T} = 2\frac{\partial}{\partial x}\left(\frac{\partial x}{\partial T}\right) \approx \frac{N_w^{\frac{1}{2}}}{2\Delta T}. \tag{17}$$

A number of assumptions and approximations were needed to arrive at Eq. (17). A preliminary set of calculations using Eq. (11) with s equal to the inverse of $N_w^{\frac{1}{2}}/2\Delta T$ from Eq. (17) resulted in an average value of divergence D unrealistically small (5×10^{-4} s^{-1}). The calculations also indicated that because D was the small difference between the two relatively large values [see Eq. (11)], a small bias in s could drastically change the results for divergence. Therefore, the parameter s is defined as

$$s = \alpha\frac{2\Delta T}{N_w^{\frac{1}{2}}}, \tag{18}$$

where α is a constant to be determined empirically through an analysis of Eq. (11).

The value for α was determined in the following manner. The terms $N^{-1}dN/dt$ and dT/dt in Eq. (11) could both be calculated from the satellite data in a limited number of cases. Thirty-eight sets of variables were available for 19 elements. The parameter s could also be calculated for these 38 cases from Eq. (18), at least in terms of α. The average divergence $\bar{D}$ in Eq. (11) was calculated using the continuity equation for an incompressible fluid,

$$\bar{D} = -\frac{\partial \bar{w}}{\partial z}, \tag{19}$$

and a composite w profile derived from the 38 cases using Eq. (1). The resulting $\bar{D}$ is 0.5×10^{-3} s^{-1}. The mean value for $N^{-1}dN/dt$ over the 38 cases is 3.4×10^{-3} s^{-1}. For $s^{-1}dT/dt$ the average magnitude is $(3.4\times10^{-3})\alpha$, where α is the proportionality constant from Eq. (18). Using these average values to substitute into Eq. (11) results in

$$0.5 = 3.4 - \frac{3.4}{\alpha}, \tag{20}$$

or

$$\alpha = \frac{3.4}{2.9} = 1.2. \tag{21}$$

With α set equal to 1.2, the shape parameter s can be calculated from satellite data from Eq. (18) in some cases. Fig. 14 displays the distribution of s as a function of temperature for the 38 cases. The values span an order of magnitude with smaller values (flatter tops) predominating at lower temperatures (higher heights), as expected. The fitted curve was determined by first dividing the data set into five approximately interval-equal groups based on temperature. Average s and temperature were calculated for each temperature interval and the linear regression curve was based on these five average points. This regression approach was adopted to reduce the effect of the relatively large number of points in the center region of the temperature range.

d. Calculation of vertical velocity

In Eq. (10), divergence is small compared to the other two terms, and to a rough approximation can be ignored to allow a direct relation between w and $N^{-1}dN/dt$. However, the effects of divergence in Eq. (10) can be partially taken into account by noting that, in the 38 cases previously examined where we estimate all of the terms in Eq. (10), D and $N^{-1}dN/dt$ appeared to be correlated. This is reasonable since D is the horizontal expansion rate of a physical area and $N^{-1}dN/dt$ is the expansion rate of area within T_{BB} isotherms. Thus D can be approximated by

$$D \approx \beta\frac{1}{N}\frac{dN}{dt}, \tag{22}$$

where β is set at 0.15, based on the average values for D and $N^{-1}dN/dt$ for the 38 cases. Substituting Eq. (22) into Eq. (10) and solving for w results in

$$w = \frac{0.85s}{(-\partial T/\partial z)}\frac{1}{N}\frac{dN}{dt}. \tag{23}$$

Eq. (23) provides a quick method of estimating vertical velocity even where there are no rapid decreases in temperature, or where an element can only be viewed for a short time.

Calculations of w using Eq. (23) are made for each of the 39 elements at particular temperatures. A lapse rate midway between a smoothed ambient sounding and a constant (9 K km^{-1}) moist adiabatic rate is used. The s values are selected from the fitted curve in Fig. 14 for the various temperatures. For temperatures $<T_{min}$ a w of zero is assigned. The calculated w values are then averaged for the severe and non-severe elements and displayed in Fig. 15. The curves are fitted subjectively.

The mean w for the severe elements is over 2.3 m s^{-1} at 226 K or 10 km and decreases rapidly with height. The non-severe elements have nearly a factor of 2 smaller vertical velocity values. Divergence values calculated for the two curves from 10–12 km using the continuity equation are 1.0×10^{-3} and 5.0×10^{-4} s^{-1} for the severe and non-severe elements, respectively.

These vertical velocity and divergence estimates are reasonable when it is recognized they are applicable

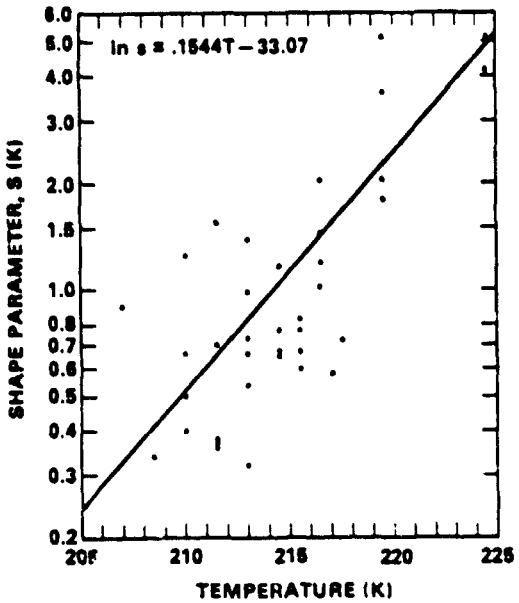


FIG. 14. Cloud top shape parameter s as calculated using Eq. (18) as a function of temperature. The 38 points are from 19 elements.

to areas ~15 km in a side. Vertical velocity estimates are very scale-dependent, and it is difficult to find values applicable on a scale equal to that determined from the satellite data. Cloud top vertical velocities estimated from aircraft observations and radar data provide at least a starting point. Observing penetrating or overshooting domes from aircraft produces vertical velocity estimates of the order of 1 m s^{-1}. For example, Umenhofer (1975) calculates the ascent of an overshooting dome to be approximately 2 m s^{-1} over a 5 min period before the Ethel, Texas, tornado of 14 May 1974 and Shenk (1974) shows another example with a w of about 3 m s^{-1}. According to Shenk the mean diameter of the overshooting domes is approximately 5.5 km which corresponds to a circular area of 24 km^2. This area is an order of magnitude smaller than the appropriate satellite area of 15^2 or 225 km^2. An example of w determined from radar top ascent can be determined from Burgess and Lemon (1976; see their Fig. 5.6). Prior to the Union City, Oklahoma, tornado, the supercell top ascended at ~3 m s^{-1}. This velocity is applicable to an area about 1–2 km on a side.

From the preceding comparisons it is obvious that the vertical velocities at cloud top derived in this paper are smaller than those derived from other sources, but because they are applicable to an area much larger, they are still reasonable. Closer comparison between the satellite information and radar, aircraft or other observations is necessary for full validation.

8. Conclusions

Digital infrared data from a geosynchronous satellite (SMS 2) have been used to study thunderstorm growth rates on 6 May 1975 in relation to the occurrence of tornadoes and other severe weather. Areal expansions of cold areas, delineated by blackbody temperature isotherms, are shown to be useful in monitoring thunderstorm growth rates, even when there is minimal decrease with time of the lowest cloud top temperature.

An examination of five cloud elements having eight tornadoes clearly associated with them indicates that in seven of the eight cases the first report of the tornado took place during, or just after, a rapid expansion of cold areas indicating rapid ascent to the thunderstorm top on the scale observed by the satellite. The rate of ascent is estimated to be approximately 0.4 m s^{-1}, averaged over an area of about 15 km on a side.

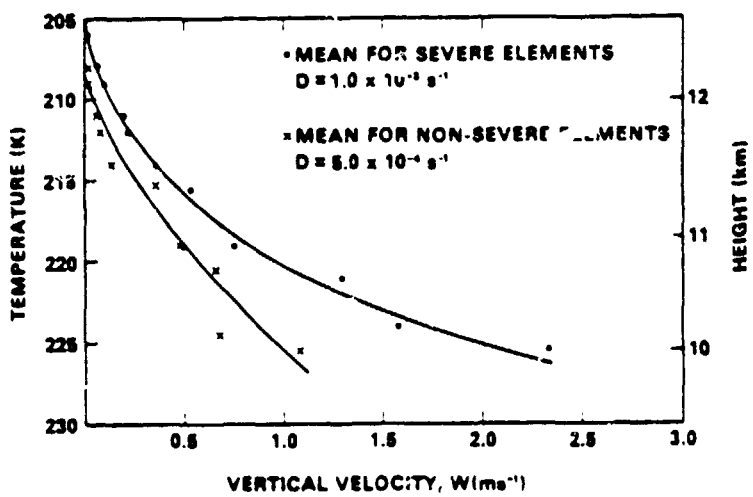


FIG. 15. Mean vertical velocity and divergence for severe and non-severe thunderstorm elements.

Thirty-nine thunderstorm elements in the analysis area were compared in relation to various parameters derived from the satellite data and the occurrence of severe weather. Although the ground truth used for defining the occurrence of severe weather may have errors, in this study two satellite-based variables are correlated with tornadoes and hail. These variables are the minimum cloud top temperature during the lifetime of the element and the logarithmic rate of increase of cloud top cold area.

The severe thunderstorm elements tend to be colder in terms of blackbody temperature and grow more rapidly than non-severe elements. For both variables the frequency distributions for severe and for non-severe elements were shown to be significantly different. A discriminant analysis using both variables showed that on one side of the discrimination boundary four out of 23 cases had severe weather, while on the cold temperature, fast growth rate side 11 out of 16 elements had reported severe weather. Therefore, based on this limited, dependent sample, these parameters, which are indicative of intense convection, are correlated with severe weather on the ground. A potential warning lead time of 30 min was estimated by comparing times of severe weather reports and times the satellite information met the derived criteria.

Equations were derived relating the areal expansion of blackbody temperature isotherms to vertical velocity and divergence. Mean vertical velocity profiles were calculated for severe and non-severe elements, with the results showing that the severe elements have a mean vertical velocity twice as large as the non-severe elements. The technique developed allows vertical velocities and divergence to be calculated from the satellite data on a horizontal spatial scale approaching that of individual thunderstorms. Information such as this will be useful in future work on tornado cloud top comparisons, further comparison of severe and non-severe elements, and for comparison with and verification of various types of cumulus models.

In summary, convection intensity information derived with short interval (~5 min) infrared digital data from a geosynchronous satellite can be statistically related to severe weather reports on the ground. Reasonable vertical velocity and divergence estimates can also be made from the data. The satellite data and the techniques developed have the potential to be a part of a severe thunderstorm detection and monitoring system.

REFERENCES

Adler, R. F., and D. D. Fenn, 1976: Thunderstorm monitoring from a geosynchronous satellite. *Preprints 7th Conf. Aerospace and Aeronautical Meteorology*. Melbourne, Amer. Meteor. Soc., 307–311.

——, and ——, 1978: Thunderstorm vertical velocities estimated from satellite data. *Preprints Conf. Weather Forecasting and Analysis and Aviation Meteorology*, Silver Spring, Amer. Meteor. Soc., 206–210.

Arn, R. M., 1975: Anvil area and brightness characteristics as seen from geosynchronous satellites. M.S. thesis, Dept. of Atmospheric Science, Colorado State University.

Billingsley, J. B., 1976: Interactive image processing for meteorological applications at NASA/Goddard Space Flight Center. *Preprints 7th Conf. Aerospace and Aeronautical Meteorology and Symposium on Remote Sensing from Satellites*, Melbourne, Amer. Meteor. Soc., 268–275.

Burgess, D. W., and L. R. Lemon, 1976: Union City storm history. Union City, Oklahoma Tornado of 24 May 1973. NOAA Tech. Memo. ERL NSSL-80, 35–51.

Donaldson, R. J., R. M. Dyer and M. J. Kraus, 1975: An objective evaluator of techniques for predicting severe weather events. *Preprints Ninth Conf. Severe Local Storms*, Norman, Amer. Meteor. Soc., 321–326.

Fujita, T. T., 1973: Proposed mechanism of tornado formation from rotating thunderstorms. *Preprints Eighth Conf. Severe Local Storms*, Amer. Meteor. Soc., Denver, 191–196.

Goddard, B., and B. Remondi, 1975: Diagnostic tools for quality assessment of VISSR data. Central processing and analysis of geostationary satellite data. NOAA Tech. Memo. NESS 64, 48–58.

Lemon, L. R., D. W. Burgess and R. A. Brown, 1978: Tornadic storm airflow and morphology derived from single-doppler radar measurements. *Mon. Wea. Rev.*, 106, 48–61.

Negri, A. J., D. W. Reynolds and R. A. Maddox, 1976: Measurements of cumulunimbus clouds using quantitative satellite and radar data. *Preprints Seventh Conf. Aerospace and Aeronautical Meteorology*, Melbourne, Amer. Meteor. Soc., 119–124.

Panofsky, H. A., and G. W. Brier, 1963: *Some Applications of Statistics to Meteorology*. The Pennsylvania State University Press, 224 pp.

Pearl, W., 1974: Characteristics of anvil-top associated with the Poplar Bluff tornado of May 7, 1973. SMRP Res. Pap. No. 119, The University of Chicago, 12 pp.

Pryor, S. P., 1978: Measurement of thunderstorm cloud-top parameters using high frequency satellite imagery. M.S. thesis, Dept. of Atmospheric Science, Colorado State University, 91 pp.

Purdom, J. F. W., 1971: Satellite imagery and severe weather warnings. *Preprints 7th Conf. Severe Local Storms*, Amer. Meteor. Soc., Kansas City, 120–137.

——, 1974: Some uses of high-resolution GOES imagery in the mesoscale forecasting of convection and its behavior. *Mon. Wea. Rev.*, 104, 1474–1483.

Shenk, W. E., 1974: Cloud top height variability of strong convective cells. *J. Appl. Meteor.*, 13, 917–991.

Sikdar, D. N., V. E. Suomi and C. E. Anderson, 1970: Convective transport of mass and energy in severe storms over the United States—an estimate from a geostationary altitude. *Tellus*, 22, 521–532.

Umenhofer, T. A., 1975: Overshooting top behavior of three tornado-producing thunderstorms. *Preprints Ninth Conf. Severe Local Storms*, Norman, Amer. Meteor. Soc., 96–104.

Yuen, T.-W., 1971: Observations of severe convective storms from SMS-1 satellite. Studies of Soundings and Imaging Measurements. Final Scientific Report, Contract NAS5-21798, University of Wisconsin, 1–64.

Paper 39

From the 13th International Symposium on Remote Sensing of Environment, Environmental Research Institute of Michigan, Ann Arbor, Michigan, April 23-27, 1979.

[D-14]

DIGITAL CLOUD STEREOGRAPHY FROM GEOSTATIONARY ORBIT

J. T. Dalton, M. L. desJardins,
A. F. Hasler and R. A. Minzner

NASA/Goddard Space Flight Center
Greenbelt, Maryland, USA

SUMMARY

It has been demonstrated that geostationary satellite imagery provides an effective means of extracting two-dimensional cloud motion wind measurements at frequent intervals over large areas. The addition of cloud height information can yield wind fields in three dimensions. This paper discusses the methodology of extracting three-dimensional cloud motion measurements from stereo digital imagery acquired at geostationary orbit. Accuracy of the technique is considered, and future enhancements and applications are explored.

Paper 40

NASA Technical Memorandum 79645 September 1978

PREDICTING TROPICAL CYCLONE INTENSITY USING SATELLITE MEASURED EQUIVALENT BLACKBODY TEMPERATURES OF CLOUD TOPS

R. Cecil Gentry*
Edward Rodgers**
Joseph Steranka*
William E. Shenk**

ABSTRACT

A regression technique has been developed to forecast 24-hour changes of the maximum winds for weak (maximum winds $\leq$ 65 kt) and strong (maximum winds $>$ 65 kt) tropical cyclones by utilizing satellite measured equivalent blackbody temperatures (T_{BB}) around the storm alone and together with the changes in maximum winds during the preceding 24 hours and the current maximum winds. Independent testing of these regression equations showed that the mean errors made by the equations are lower than the errors in forecasts made by the persistence techniques.

*General Electric Space Division (MATSCO), Beltsville, Maryland 20705
**Laboratory for Atmospheric Science (GLAS) NASA/GSFC, Greenbelt, Maryland 20771

C-5

Paper 41

NASA Conference Publication 2076 – Fourth NASA Weather and Climate Program Science Review, January 24-25, 1979, NASA/GSFC, Greenbelt, MD.

Paper No. 12

FOUR DIMENSIONAL OBSERVATIONS OF CLOUDS FROM GEOSYNCHRONOUS ORBIT USING STEREO DISPLAY AND MEASUREMENT TECHNIQUES ON AN INTERACTIVE INFORMATION PROCESSING SYSTEM

A. F. Hasler, M. deJardins, W. E. Shenk, *Goddard Space Flight Center, Greenbelt, Maryland*

ABSTRACT

Simultaneous Geosynchronous Operational Environmental Satellite (GOES) 1 km resolution visible image pairs can provide quantitative three dimensional measurements of clouds. These data have great potential for severe storms research and as a basic parameter measurement source for other areas of meteorology (e.g. climate). These stereo cloud height measurements are not subject to the errors and ambiguities caused by unknown cloud emissivity and temperature profiles that are associated with infrared techniques. Previous work by Minzner et al. (1978a) has demonstrated the validity of stereo measurements from geostationary satellites using techniques based on conventional analogue stereography. This effort describes the display and measurement of stereo data using digital processing techniques.

Computer remapping of digital GOES image pairs allows the interactive display (on the Atmospheric and Oceanographic Information Processing System, AOIPS) of time sequences of stereo images a true four dimensional representation of cloud structures. A description of similar work at the University of Wisconsin is given by Bryson (1978). Interactive manual and semi-automatic height measuring techniques have also been developed on AOIPS. Capabilities under development include 1) measurement of multilevel wind fields with accurate height assignment; 2) estimation of thunderstorm intensity from horizontal and vertical cloud growth rates, and 3) measurement of four dimensional cloud structure for comparison with numerical models and radar observations. Stereo height verification and error analyses were conducted using computer crosscorrelation on AOIPS. Accuracies of ±0.5 km appear to be possible for geographical features and clouds with precisions (relative accuracies) approaching ±0.1 km.

Paper 42

Reprinted from JOURNAL OF APPLIED METEOROLOGY, Vol. 18, No. 11, November 1979
American Meteorological Society
Printed in U. S. A.

In Situ Aircraft Verification of the Quality of Satellite Cloud Winds over Oceanic Regions

A. F. HASLER, W. C. SKILLMAN AND W. E. SHENK

Laboratory for Atmospheric Sciences, Goddard Space Flight Center, NASA, Greenbelt, MD 20771

J. STERANKA

GEIMATSCO, Beltsville, MD 20705

(Manuscript received 23 March 1979, in final form 4 August 1979)

ABSTRACT

A 5-year aircraft experiment to verify the quality of satellite cloud winds over oceans using *in situ* aircraft Inertial Navigation System wind measurements has been completed. The final results show that satellite measured cumulus cloud motions V_{cloud} are very good estimators of the cloud-base wind V_{CBW} (900–950 mb) for trade wind and subtropical high regions. The average magnitude of the vector differences between the cloud motion and the cloud-base wind ranged from 0.9 to 1.7 m s^{-1} [0.9 m s^{-1} $\leq |\overrightarrow{V_{cloud} - V_{CBW}}| \leq$ 1.7 m s^{-1}]. For cumulus clouds near frontal regions, the cloud motion agreed best with the mean cloud layer wind V_{MCLW} [$|\overrightarrow{V_{cloud} - V_{MCLW}}|$ = 2.3 m s^{-1}]. For a very limited sample, cirrus cloud motions also most closely followed the mean wind in the cloud layer [$|\overrightarrow{V_{cloud} - V_{MCLW}}|$ = 1.7 m s^{-1}].

1. Introduction

A global system of five geosynchronous satellites is now in place. Currently it consists of the United States SMS/GOES satellites located at 75°W, 135°W and 60°E longitude; the European Meteosat at 0° longitude and the Japanese GMS at 140°E longitude. This system is likely to be permanent except that the U.S satellite at 60°E will be replaced by the Indian Insat in 1981. One of the prime objectives of this system is to provide winds from cloud motions around the globe at low and middle latitudes for improved numerical weather forecasting. Many researchers, including Fujita *et al.* (1969), Hasler (1972), Smith and Hasler (1976), Suchman *et al.* (1977), Rodgers *et al.* (1979) and Peslen (1979), have shown that wind fields from satellite cloud motions provide good coverage for a wide variety of atmospheric phenomena. They have also demonstrated that cloud wind analyses give reasonable descriptions of phenomena which are consistent with accepted theories.

The American, European and Japanese meteorological agencies are providing satellite-derived cloud winds at least twice a day on an operational basis.

In view of this wide use of satellite-derived cloud winds and the associated high expenditure of resources it is imperative that their quality be validated. Except for the recent research work by Rodgers *et al.* (1979) and Peslen (1979) and others with special short time interval images almost all satellite cloud winds for research and operations have been determined from images at 30 min intervals, with horizontal resolutions ranging from 2 to 8 km.

According to Malkus (1949) small cumulus clouds would move at the ambient wind velocity in the absence of vertical shear of the horizontal wind. When vertical shear is present Malkus found that the cumulus cloud will move with a velocity that is primarily a function of the cloud-base wind and the magnitude of the shear. However, the 30 min geosynchronous satellite observation interval would not temporally resolve the individual cumulus cloud elements. Therefore, with the 30 min interval, the satellite is only able to follow the history of an ensemble of cumulus cells, a small mesoscale feature with a horizontal dimension of not usually less than 10 km. Telford and Wagner (1974) and in more recent unpublished work have concluded that the growth and decay of small cumuli are linked with the air movement at the heat and moisture source below the cloud. Thus it is reasonable to expect satellite-observed cumulus to move closest to the cloud-base wind or the wind below cloud base. Hasler *et al.* (1976) have presented high-resolution aerial photography taken as frequently as once every 7 min which show that the lifetimes of individual cells are short, but that cumulus ensembles can maintain a recognizable pattern for well over an hour. The cross section in Fig. 1 illustrates conceptually the evolution of a part of a typical cumulus ensemble in a

0021-8952/79/111481-09506.25

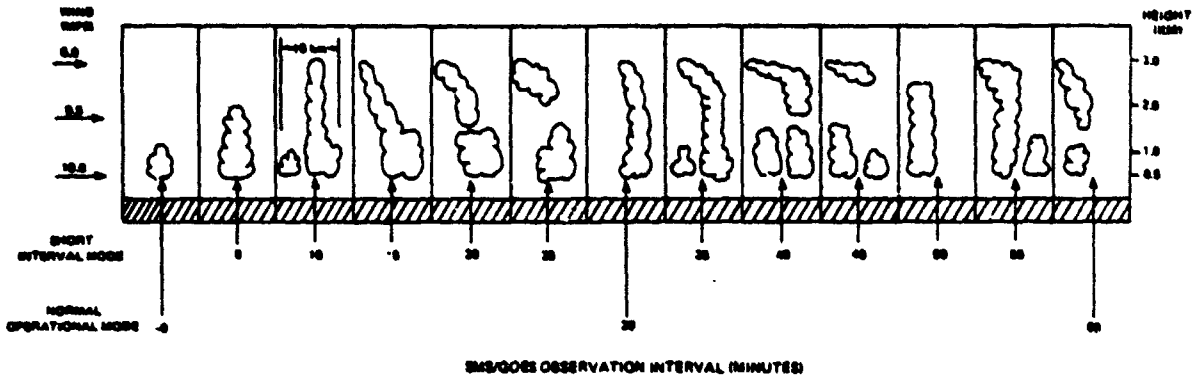


FIG. 1. Conceptual model of the motion of a typical cumulus cloud ensemble in a vertical wind shear environment as viewed by a geosynchronous satellite.

vertical wind shear environment. Individual cumulus cells continue to be generated, probably at the location of maximum small-scale convergence near the original center of activity which is being transported with the wind at cloud base. The tops of the towers are repeatedly separated from the convection center because of the vertical wind shear.

Following the arguments of Malkus, detached passive cirrus cloud elements would be expected to move with the ambient wind in the absence of substantial vertical motion. Active cirrus clouds with upward vertical motion should, like the cumulus, have velocities which are a function of the cloud-base wind and the vertical shear. For cirrus, however, cloud lifetimes can often be resolved by the 30 min satellite observations.

There are some clouds which should not have a useful relationship to the ambient flow. Orographic clouds tend to be stationary and clouds caused by gravity waves tend to move with the wave phase velocity and neither would be good estimators of the ambient wind. Studies using radar (Battan, 1973) show that cumulonimbus clouds move with a velocity which is a function of the integrated wind through most of the troposphere.

Comparisons of satellite cloud motions and wind evaluations using rawinsondes have been performed by Fujita *et al.* (1969), Hubert and Whitney (1971), Hasler (1972), Fujita *et al.* (1975), Bauer (1976), Suchman and Martin (1976) and Hubert.[1] Hubert's latest study best characterizes this type of evaluation. He finds, for low cloud motions derived from the NOAA/NESS operational system, that the median magnitude of the vector difference from 900 mb rawinsonde winds is 2.6 m s^{-1}, while 68% of the differences are less than 4.0 m s^{-1}. For cirrus cloud motions the median magnitude of the vector difference from rawinsonde winds at the assigned level is 5.7 m s^{-1} and 68% of the differences are less than 8.0 m s^{-1}. This type of comparison is limited by large time and space differences between the observations. In Hubert's work, for example, the time differences are up to 3 h, and the horizontal space differences are up to 300 km. Errors in height assignment, particularly for cirrus clouds, are likely to account for a large portion of the differences.

Telford and Wagner (1974) have done limited comparisons of aircraft-measured cloud motions with *in situ* aircraft winds over land. For three cumulus clouds Wagner and Telford's data from more recent unpublished work show that the magnitude of the vector difference between the cloud motion and the wind below cloud base was ~1.0 m s^{-1}.

The technique used in this study was a comparison of cloud motions measured by satellite and aircraft with aircraft wind measurements that were coincident in time and space (Hasler *et al.*, 1976, 1977). The results from the 5-year experiment are for undisturbed to moderately disturbed oceanic weather regimes. The experiment was conducted in five phases where a total of 77 cloud motion measurements were compared with the ambient winds. The preliminary results from the first three phases have been reported by Hasler *et al.* (1976, 1977). The locations and meteorological conditions of the five phases of the experiment are summarized in Fig. 2 and Table 1. In Phase I, a small sample of cumulus clouds (6) and one cirrus cloud were tracked by the NCAR Sabreliner in December 1972 in the northwest Caribbean under conditions of moderate winds and vertical wind shear.

Phase II of the experiment was flown in April 1974 in the southwest Caribbean near Panama. Nineteen low-level clouds were tracked by the NASA C-130 and NCAR Sabreliner under moderate trade wind conditions, while four additional cirrus clouds were

[1] Hubert, L. F., 1976: Wind determination from geostationary satellites. *Proc. Symp. Meteorological Observations from Space: Their contribution to the First Garp Global Experiment*. COSPAR, ICSU, National Center for Atmospheric Research, Boulder, 211–213. [Available from National Center for Atmospheric Research, Boulder CO 80302.]

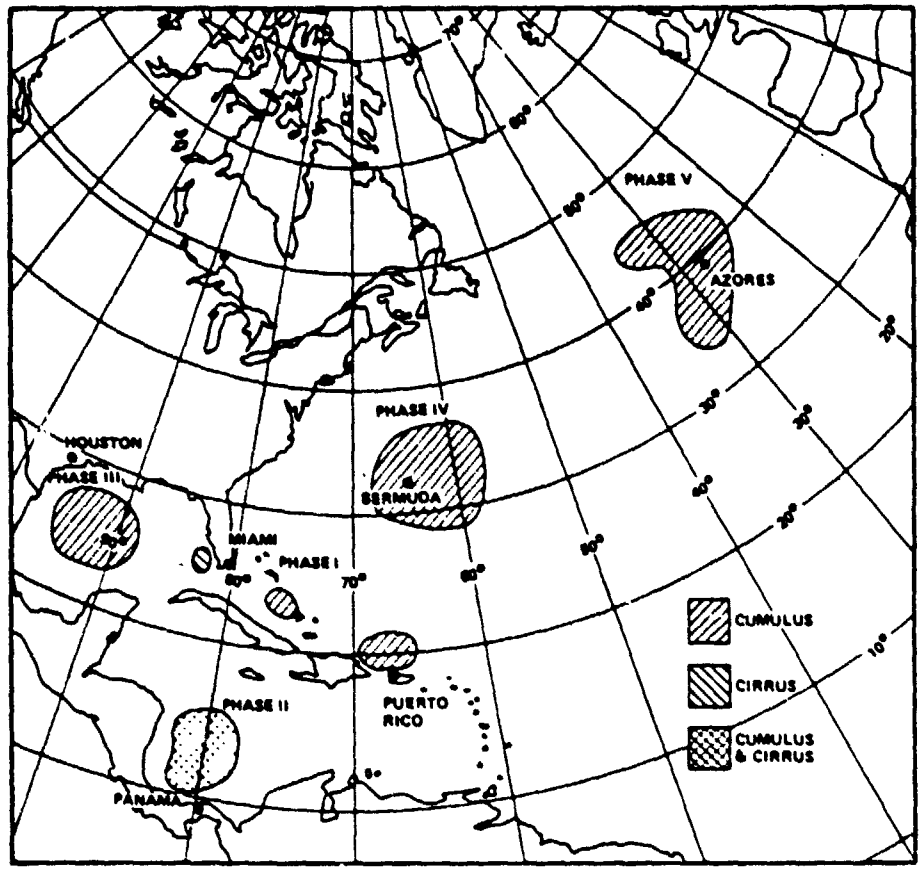


FIG. 2. Locations for the five phases of an experiment for the *in situ* verification of satellite cloud winds using aircraft.

tracked by the Sabreliner. Phase III was conducted in July 1974 in the Gulf of Mexico under light wind and shear conditions where 15 low clouds were tracked. Phase IV was accomplished in January and February 1976 in the northwestern Atlantic during high wind frontal weather situations near Bermuda which resulted in 25 additional cloud tracking cases. Phase V of the experiment tracked 11 more clouds during February 1977 in the northeastern Atlantic under the high wind suppressed convection conditions in the Azores subtropical high.

As an illustration of the type of clouds tracked in this experiment, the DMSP high-resolution (600 m) visible satellite image in Fig. 3 shows a typical field of cumulus clouds from which a wind tracer was selected in the Phase II tracking area north of Panama. Most of the clouds inside the dashed line can be seen in the 900 m resolution SMS/GOES images and a large fraction have lifetimes of 30 min or longer which was required in this experiment and would serve as good wind tracers.

2. Techniques

The basic objective of the experiment is to measure cloud motions from geosynchronous satellite image sequences and to simultaneously determine the wind field of the cloud environment with an

TABLE 1. Dates, locations and conditions for the five phases of an experiment for the *in situ* verification of cloud winds

Phase	Date	Location	Weather regime	Average low-level wind speed
I	December 1972	NW Caribbean	trade wind	moderate 12 m s^{-1}
II	April 1974	SW Caribbean	trade wind	moderate 10 m s^{-1}
III	July 1974	Gulf of Mexico	trade wind	light 5 m s^{-1}
IV	Jan.–Feb. 1976	NW Atlantic	frontal	high 17 m s^{-1}
V	February 1977	NE Atlantic	subtropical high	high 17 m s^{-1}

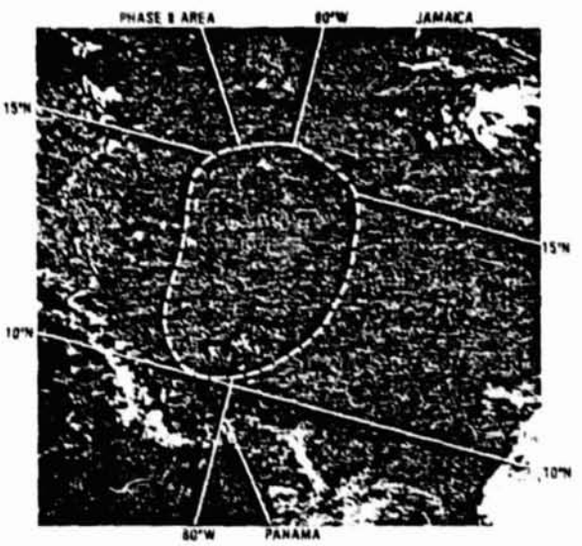


FIG. 3. High-resolution visible DMSP satellite photo of the southwest Caribbean for 1616 GMT 1 April 1974 showing typical low-level cloud tracers which were tracked in the experiment. The Phase II cloud-tracking area north of Panama is marked.

aircraft. It was necessary to find isolated clouds with no overcast above, fix their positions as a function of time and measure the wind at the cloud top, mid cloud, cloud base and for low-level cumulus clouds, in the subcloud layer (150 m). Cloud positions and wind measurements were made for at least 30 min and up to 2.5 h for each cloud. The wind measurements were derived from the Inertial Navigation System (INS) of the aircraft. At each level a straight, constant-altitude flight line was made through the cloud with at least 2 min in the clear air on each side. This was repeated on a reverse heading so that bias errors originating from air speed, ground speed, heading, track and angle of sideslip could be corrected. Grossman (1977) provides a procedure for this correction.

The Grossman technique involves the application of derived wind component equations which assume wind field homogeneity and negligible error in aircraft heading. The uncorrected wind measurements are employed to obtain the component biases along the longitudinal and lateral axes of the aircraft. These biases are then applied to the uncorrected wind in order to arrive at a corrected wind. In this investigation, recorded in-flight wind data were available along the entire length of the aircraft track from the end of a turn, through a cloud, to the beginning of a new turn and a reverse track along approximately the same path. The wind data from cloud exit to the beginning of a turn were averaged to provide an average wind for one leg and the wind data from the end of a turn to cloud entry were averaged for the average wind of the second leg. These average winds, when employed with the bias component equations, permitted the computation of component bias corrections. The resultant combination of corrections and measured wind was utilized as the *in situ* cloud level wind for the comparisons with cloud motions. Cloud positions were determined at intervals of 5–10 min using the latitude and longitude given by the aircraft INS as shown in the example in Fig. 4. These positions were used to compute the cloud motion and to locate the cloud in the satellite images. Further details on the experimental procedure using the aircraft are given by Hasler *et al.* (1976, 1977).

The satellite cloud motions were measured for the 1972 northwest Caribbean case (Phase I) by techniques described by Hasler *et al.* (1976). For the 1974, 1976, 1977 phases, measurements were made from sequences of digital SMS/GOES images on the Atmospheric and Oceanographic Information Processing System (AOIPS) at the Goddard Space Flight Center (GSFC). Descriptions of the AOIPS hardware and software systems are given by Billingsley and Hasler[2] and Bracken *et al.*[3] Cloud motions were determined on AOIPS from image sequences of at least 60 min with a partial or complete overlap in time with the aircraft observations. In most cases there was little doubt that the cloud tracked by the satellite and the aircraft were the same. In a few cases due to high-level overcast or confusion caused by nearby clouds, a similar cloud within 25 km of the aircraft location was used. According to procedures developed by Hasler and Rodgers,[4] satellite cloud velocity errors of 0.5 m s^{-1} would be expected using an image sequence of 60 min with 1 km resolution SMS/GOES images on AOIPS. The aircraft wind measurements are accurate to 1.4 m s^{-1} (Kelly and Zruber, 1973). Therefore, adding the expected error in the satellite measured cloud velocity to the expected aircraft wind error in a root-mean-square sense the expected measurement error is 1.5 m s^{-1}. The error in the aircraft measured cloud velocities is 0.6 m s^{-1} (Hasler *et al.* 1976) so the rms sum is also ~1.5 m s^{-1} for aircraft cloud velocity vs. aircraft wind comparisons. Adding the expected aircraft cloud velocity error, 0.6 m s^{-1}, to the expected

[2] Billingsley, J. B., and A. F. Hasler, 1975: Interactive image processing for meteorological applications at NASA Goddard Space Flight Center. Rep. X933-75-97, Goddard Space Flight Center, Greenbelt, MD 20771.

[3] Bracken, P. A., J. T. Dalton, J. B. Billingsley and J. J. Quann, 1977: Atmospheric and Oceanographic Information Processing System (AOIPS). System description. Rep. X933-77-148, Goddard Space Flight Center, Greenbelt, MD 20771.

[4] Hasler, A. F., and E. B. Rodgers, 1977: An error analysis of tropical cyclone divergence and vorticity fields derived from satellite cloud winds on the Atmospheric and Oceanographic Image Processing Systems (AOIPS). *Preprints 11th Tech. Conf. Hurricanes and Tropical Meteorology*, Miami, Amer. Meteor. Soc., 670–675.

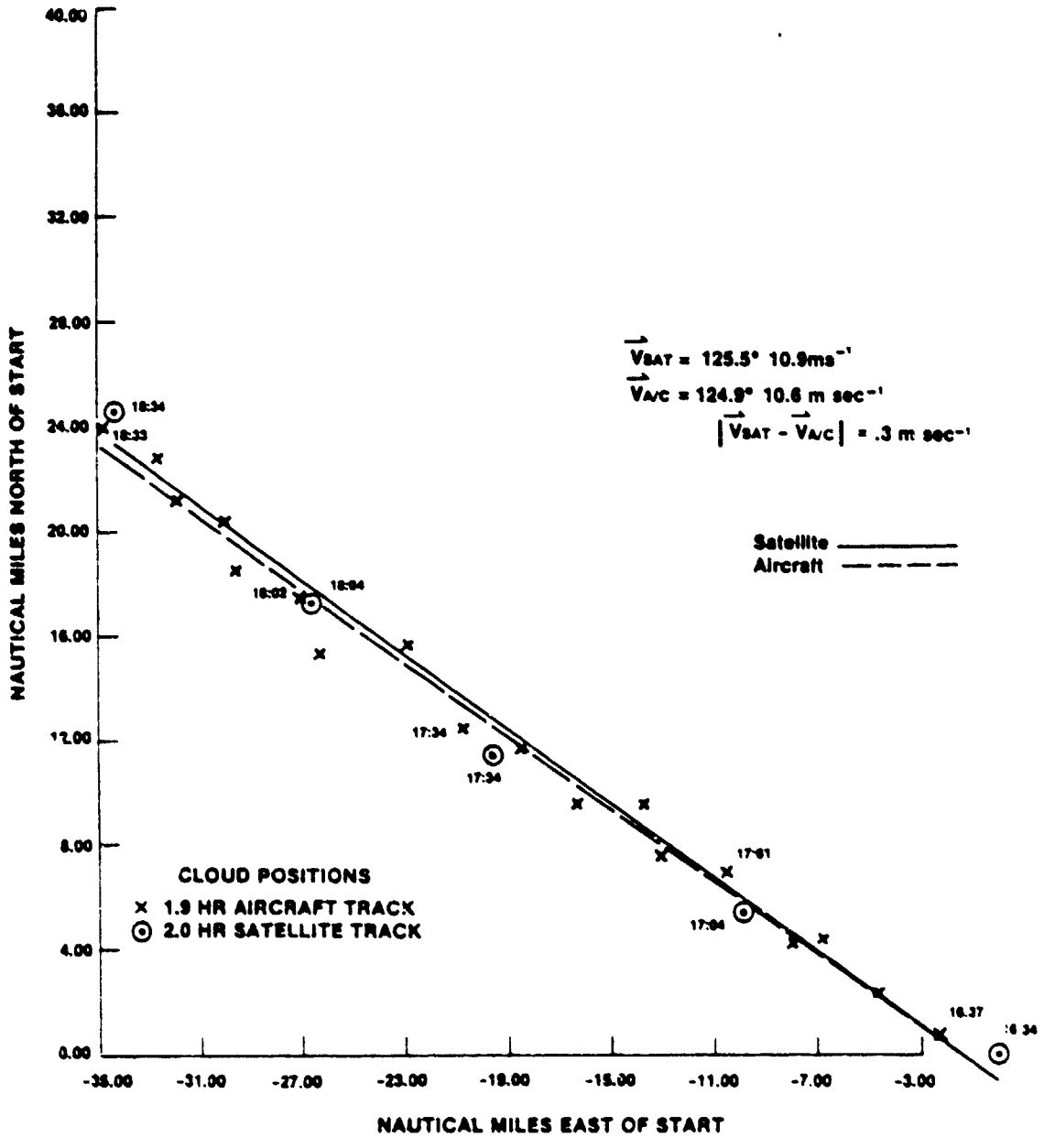


FIG. 4. Satellite versus aircraft cloud tracking in the northwest Atlantic (31°N, 67°W), 27 January 1976.

satellite cloud velocity error, 0.5 m s^{-1}, in the same manner yields a total expected measurement error of 0.8 m s^{-1} for comparisons.

3. Results

a. Aircraft measured cloud motions versus winds.

In Table 2 the final results for all five phases of the experiment are presented. For low-level cumulus in moderate trade winds (average cloud speed of 8.7 m s^{-1}) and high wind speed (average cloud speed of 17.8 m s^{-1}) subtropical high regimes, the cloud motions agree best with the winds at cloud base [1.5 m s^{-1} ≤ $|\overline{V_{cloud} - V_{CBW}}|$ ≤ 1.8 m s^{-1}]. Both the average magnitude of the vector difference [$|\overline{V_{cloud} - V_{wind}}|$] and the difference within which two-thirds of each sample were contained [$|V_{cloud} - V_{wind}|_{67\%}$] are presented. For low-level cumulus in high-speed wind regimes near fronts the cloud motions are closest to the mean wind in the cloud layer [$|\overline{V_{cloud} - V_{MCLW}}$ = 2.5 m s^{-1}]. Surprisingly, the relationship was the same for the relatively deep

TABLE 2. Cloud motion versus *in-situ* aircraft wind *

Cloud type	Oceanic weather regimes	Date	Location	Number of cloud tracks	Average cloud speed (m s^{-1})	$\lvert\overline{V_{cloud} - V_{wind}}\rvert$ / $\lvert V_{cloud} - V_{wind}\rvert_{67\%}$** (m s^{-1}) 150 m	Cloud base	Mid cloud	Cloud top	Mean in cloud layer
Low-level cumulus	Trade wind	Dec 72 Apr 74 Jul 74	NW Caribbean SW Caribbean Gulf of Mexico	40	8.7	1.6/1.9	1.5/1.5	2.9/3.4	6.2/6.8	2.8/3.2
	Subtropical high	Feb 77	NE Atlantic	6	17.8	3.5/3.7	1.8/1.9	5.4/4.6	9.6/9.1	5.1/4.8
	Frontal	Feb 76	NW Atlantic	18	15.3	3.6/3.8	3.2/3.8	2.8/3.3	5.6/6.9	2.5/2.9
Cirrus	Subtropics	Dec 72 Apr 74	W Caribbean	5	11.0	—	2.1/1.7	2.0/1.8	2.9/2.2	1.7/2.0

* All measurements were made *in-situ* by aircraft equipped with Inertial Navigation Systems.
** The 67% subscript means that two-thirds of the differences have this value or less

convective region ahead of the front and the suppressed convection behind it.

For the few cirrus cases that were tracked in the subtropics, Table 2 shows that the cloud motion agreed best with the mean wind in the cloud layer [$|\overline{V_{cloud} - V_{MCLW}}| = 1.7$ m s^{-1}], but the agreement at any of the three levels is not significantly different.

b. *Satellite measured cloud motions versus winds*

Table 3 shows cloud motion versus ambient wind comparisons where the cloud motions were measured from sequences of geosynchronous satellite visible images on AOIPS. The satellite cloud winds versus aircraft wind comparisons in Table 3 give the same general results as those in Table 2. However satellite data were not available for the 1974 southwest Caribbean cumulus and cirrus cases. For trade wind and subtropical high cumulus clouds, the motions agree best with the winds at cloud base or below [0.9 m s$^{-1} \leq |\overline{V_{cloud} - V_{CBW}}| \leq 1.7$ m s^{-1}]. The cumulus cloud motions near fronts again were closest to the mean wind in the cloud layer [$|\overline{V_{cloud} - V_{MCLW}}| = 2.3$ m s^{-1}].

c. *Satellite vs aircraft cloud motions*

Fig. 4 shows a comparison of an aircraft cloud track with a satellite cloud track where an isolated well defined cloud was tracked over nearly identical 2 h periods of time by both systems. For this case the magnitude of the vector difference was $|\overline{V_{cloud_s} - V_{cloud_a}}| = 0.3$ m s^{-1}. For the cases where aircraft and satellites tracked the same clouds the average differences were $|\overline{V_{cloud_s} - V_{cloud_a}}| = 1.4$, 1.8, 1.5 and 1.6 m s^{-1} for the 5, 16, 4 and 10 clouds tracked in the northwest Caribbean, Gulf of Mexico, northwest Atlantic and northeast Atlantic, respectively. These differences are about twice as large as the 0.8 m s^{-1} expected error which was calculated in the previous section. This indicates that some clouds must not have been properly identified in the satellite pictures or and measurement errors were larger than anticipated in some cases.

d. *Bias errors*

Table 4 gives the satellite cloud wind (satellite cloud motion) bias errors with respect to the cloud-

TABLE 3. Satellite cloud winds versus *in situ* aircraft wind measurements for oceanic cumulus clouds

Weather regimes	Date	Location	Number of cloud tracks	Average cloud speed (m s^{-1})	$\lvert\overline{V_{cloud} - V_{wind}}\rvert$ / $\lvert V_{cloud} - V_{wind}\rvert_{67\%}$ (m s^{-1}) 150 m	Cloud base	Mid cloud	Cloud top	Mean in cloud layer
Trade wind	Dec 72	NW Caribbean	6	12.2	—	0.9/1.2	3.2/3.0	5.2/6.0	2.1/3.4
Trade wind	Jul 74	Gulf of Mexico	10	5.3	1.4/1.7	1.7/1.8	3.4/2.9	5.4/5.4	2.2/3.0
Subtropical	Feb 77	NE Atlantic	7	19.0	4.3/4.1	1.7/1.8	4.8/4.3	10.7/11.5	4.8/5.3
Frontal	Feb 76	NW Atlantic	19	15.7	4.1/4.2	3.6/4.0	3.0/3.0	5.3/7.8	2.3/2.5

See footnotes to Table 2.

TABLE 4. Satellite cloud wind systematic bias errors for oceanic cumulus clouds.

		Satellite cloud motion versus cloud-base wind					Satellite cloud motion versus cloud-layer wind				
					Systematic bias (cloud − CBW)					Systematic bias (cloud − MCLW)	
Weather regimes	Location	Number of cloud tracks	$\bar{V}_{cloud}$ (m s^{-1})	$\bar{V}_{CBW}$ (m s^{-1})	Speed (m s^{-1})	Direction (deg)	Number of cloud tracks	$\bar{V}_{cloud}$ (m s^{-1})	$\bar{V}_{MCLW}$ (m s^{-1})	Speed (m s^{-1})	Direction (deg)
Trade wind	NW Caribbean	6	12.2	12.3	−0.1	0.5	4	12.6	10.8	1.8	−8.2
	Gulf of Mexico	10	5.3	5.3	0.0	5.0	10	5.3	3.9	1.3	14.8
Subtropical high	NE Atlantic	7	19.0	18.0	1.0	1.6	7	19.0	23.0	−4.0	3.7
Frontal	NW Atlantic	19	15.7	16.2	−0.5	5.1	19	15.7	16.7	−1.0	−1.3

base wind and also the mean wind in the cloud layer (cloud-layer wind). For the cloud wind biases with respect to the cloud-base wind, all speed biases are 1 m s^{-1} or less. Direction biases are also small for the trade wind northwest Caribbean case and the subtropical high cases. For the Gulf of Mexico trade wind case, the 5.0° systematic direction bias error contributes only slightly to the mean magnitude of the vector difference [$|\overline{\mathbf{V}_{cloud} - \mathbf{V}_{CBW}}| = 1.7$ m s^{-1}, from Table 3] because of the low average wind speed of 5.3 m s^{-1}. For the frontal case a nearly identical direction bias error of 5.1° contributes much more to the large vector difference [$|\overline{\mathbf{V}_{cloud} - \mathbf{V}_{CBW}}| = 3.6$ m s^{-1}] because of the high average wind speed of 16.2 m s^{-1}. The cloud wind bias errors with respect to the mean wind in the cloud layer are small only for the frontal case (−1.0 m s^{-1} in speed and −1.3° in direction). Because of the high average wind speed in the frontal case the small direction bias error contributes substantially to the vector difference of $|\overline{\mathbf{V}_{cloud} - \mathbf{V}_{MCLW}}| = 2.3$ m s^{-1} between the cloud wind and the cloud layer wind. In the frontal case the cloud wind does not agree with the mean wind in the cloud layer as well as the cloud winds for the other phases agree with the cloud base winds (see Table 3). Therefore the effect of systematic bias error removal was evaluated for this case. From Table 5 it is evident that systematic error removal results in only a very small improvement in the average magnitude of the vectors difference for the mean in the cloud layer, from 2.3 to 2.2 m s^{-1}. However, the removal of large systematic differences in direction would give improvements of 4.1 to 2.6 m s^{-1} and 3.6 to 2.9 m s^{-1} for the 150 m and cloud-base levels, respectively. Therefore, if the cloud winds are used to estimate the mean wind in the cloud layer there would be little purpose in removing systematic errors; but for cloud-base or subcloud-layer wind estimation, removal of the bias errors would be advantageous.

It was also determined that the magnitude of the vector difference between cloud motion and the cloud-base wind is not highly correlated to either the wind speed or the vertical shear for the trade wind and subtropical high cases.

e. Cloud wind height assignment

Satellite cloud winds for oceanic trade wind and subtropical high regions may be assigned to the cloud-base altitude. There is no reliable way of measuring cloud-base height from geosynchronous satellites since the cloud bases are hidden from view by the upper part of the cloud so the best method for low-level cloud wind height assignment is to use climatology, aircraft reports, or deduce them from soundings where available. Cloud-base altitude statistics for the entire experiment are given in

TABLE 5. Satellite cloud winds versus *in situ* aircraft winds, systematic bias error removed* for western Atlantic (32°N) January–February 1976

	150 m	Cloud base	Mid cloud	Cloud top	Mean in cloud layer
Before removal of systematic differences $\overline{\mathbf{V}_{cloud} - \mathbf{V}_{wind}}$ (m s^{-1})	4.1	3.6	3.0	5.3	2.3
Systematic speed differences ΔS (m s^{-1})	−1.2	.5	.8	2.5	1.0
Systematic direction difference ΔD (deg)	−8.1	−5.1	3.5	6.5	1.3
Systematic differences removed $\overline{\mathbf{V}_{cloud} - \mathbf{V}_{wind}}$ (m s^{-1})	2.6	2.9	2.9	4.8	2.2

* Statistics are based on 19 cases.

TABLE 6. Cloud-base and cloud-top height statistics based on 64 cumulus and five cirrus clouds.

Cloud type	Weather regimes	Location	Cloud bases (mb)				Cloud tops (mb)			
			Mean height ($\bar{h}$)	Standard deviation (σ_h)	Minimum height (h_{min})	Maximum height (h_{max})	Mean height ($\bar{h}$)	Standard deviation (σ_h)	Minimum height (h_{min})	Maximum height (h_{max})
Low-level cumulus	Trade winds	NW Caribbean	946	11.5	960	930	721	55.6	810	205
		SW Caribbean	943	7.5	956	932	671	86.6	797	565
		Gulf of Mexico	939	5.8	942	925	492	134.4	753	301
	Subtropical high	NE Atlantic	927	7.1	974	901	753	110.7	901	558
	Frontal (N.W. Atlantic)	Prefrontal	938	23.1	977	898	613	96.8	736	416
		Postfrontal	915	15.5	944	898	757	38.9	827	716
		Combined frontal	931	23.6	977	898	659	102.7	827	416
	Combined low level		936	19.0	977	898			901	205
Cirrus		W Caribbean	275	62.2	345	197	219	64.4	331	171

Table 6. The average low-level cloud-base altitude for the experiment was 936 mb with a standard deviation of only 19 mb. There was a tendency for the low-latitude cloud bases to be lower ($\sim\bar{h}$ = 940 mb) and more uniform ($\sim\sigma_h$ = 10 mb) than the higher latitude bases ($\sim\bar{h}$ = 930 mb, $\sim\sigma_h$ = 25 mb). However, the total range of the low-level cloud bases was only from 977 to 898 mb. Since the low-level cumulus cloud bases are very uniform in altitude, assignment of the cloud winds to 950 or 900 mb should be sufficiently accurate for most applications.

In frontal regions cumulus cloud winds may be assigned to the middle of the cloud layer. This can be done by measuring the cloud-top altitude by the Mosher's combined infrared and visible method (Suomi[5]) or stereo techniques (Minzner *et al.*, 1978; Hasler *et al.*[6]) and using the cloud base of ~930 mb from Table 6 to calculate the mid-cloud level.

Cirrus cloud winds should be also assigned to the mid-cloud level. Great care must be taken not to underestimate the altitude of the cirrus cloud tops from infrared measurements, but stereo heights show considerable promise for eliminating this problem. It may be best to assign cirrus cloud winds to the cloud-top altitude, because the data show little difference between the various levels and it is difficult to make a good estimate of the cloud-base height.

4. Summary and conclusions

For oceanic trade wind cumulus and cumulus in oceanic subtropical high regions, satellite cloud motions estimate the wind at cloud base at approximately the limit of the instrumental accuracy possible from this experiment [0.9 m s^{-1} ≤ $|\overline{V_{cloud} - V_{CBW}}|$ ≤ 1.7 m s^{-1}] with no significant bias errors. For oceanic cumulus clouds near fronts, agreement is best with the mean wind in the cloud layer [$|\overline{V_{cloud} - V_{MCLW}}|$ = 2.3 m s^{-1}]. The differences are larger in this case, but removal of systematic errors produces no significant improvement. For high-level cirrus cloud motions measured by aircraft, the comparison agreement is also best with the mean wind in the cloud layer [$|\overline{V_{cloud} - V_{MCLW}}|$ = 1.7 m s^{-1}], but is not significantly better than the agreement with the cloud-base or top wind. It is concluded that for most equatorial through mid-latitude ocean areas of the world, satellite cloud motions can be used to estimate the low-level (cloud base) winds with high accuracy. There is no reliable method yet demonstrated of estimating cloud-base altitudes from geosynchronous satellite orbit, but low-level cumulus cloud bases are very uniform in height and can be determined within a few tens of millibars from climatology and/or from relatively widely spaced surface station reports. The 64 cumulus clouds measured in this experiment had an average cloud-base height of $\bar{h}$ = 936 mb with a standard deviation of only σ_h = 19 mb.

In frontal regions cumulus cloud-top altitudes should be determined from infrared or stereo measurements and the cloud wind assigned to the mid-cloud level. Cirrus cloud winds should also be assigned to the mid-cloud level if it can be determined, but assignment to the cloud top level is probably satisfactory.

Satellite cloud motions can be excellent estimators of the wind for carefully selected tracers which are not affected by gravity waves or orography. Proper height assignment of the cloud winds is also extremely important and has probably contributed most to poor wind estimation in the past (e.g., the case of thin cirrus cloud motions assigned to too low a level).

[5] Suomi, V. E., 1975: Man-computer Interactive Data Access System (McIDAS). Final Report Contract NAS5-23296, University of Wisconsin, Madison. [Available from Space Science and Engineering Center, University of Wisconsin, Madison.]

[6] Hasler, A. F., M. des Jardins and W. E. Shenk, 1979: Four-dimensional observations of clouds from geosynchronous orbit using stereo display and measurement techniques on an interactive information processing system. *Fourth National Aeronautics and Space Administration Weather and Climate Review*, Greenbelt, NASA, 67–72. [Available from Goddard Space Flight Center, Greenbelt, MD 20771.]

5. Future work

It is still necessary to obtain more data for cumulus clouds in other oceans, over land and under disturbed conditions to more fully assess cloud motion wind relationships. The sample size for cirrus clouds needs to be increased and comparisons made in high wind shear situations (e.g., jet streams). There is potential for better comparisons with rawinsondes if geosynchronous satellite stereo observations become available in regions where cloud winds can be measured coincidently in time and space with the rawinsondes. In the immediate future the experiment will concentrate on *in situ* aircraft verification under disturbed conditions, particularly over land for the antecedent conditions for severe local storms.

Acknowledgments. The authors would like to express their gratitude to the aircraft support teams from Johnson Space Center and the National Center for Atmospheric Research. Their positive attitudes toward unusual flight operations and diligence in processing of the data made this experiment possible. Ray Minzner, Gerard Szejwach and Edward Pearl assisted in taking observations and their efforts are appreciated. Robert Grossman was most helpful with supplying information and answering questions about the use of his aircraft wind bias error removal technique.

REFERENCES

Battan, L. J., 1973: *Radar Observation of the Atmosphere*. The University of Chicago Press, 324 pp. (see p. 209).

Bauer, K. G., 1976: A comparison of cloud motion winds with coinciding radiosonde winds. *Mon. Wea. Rev.*, 104, 922–931.

Fujita, T. T., K. Watanabe and T. Izawa, 1969: Formation and structure of equatorial flows, derived by ATS-1 photographs. *J. Appl. Meteor.*, 8, 649–667.

——, E. W. Pearl and W. E. Shenk, 1975: Satellite-tracked cumulus velocities. *J. Appl. Meteor.*, 14, 407–413.

Grossman, R. L., 1977: A procedure for the correction of biases in winds measured from aircraft. *J. Appl. Meteor.*, 16, 654–658.

Hasler, A. F., 1972: Properties of tropical cloud clusters determined from geostationary satellite pictures. Ph.D. thesis, University of Wisconsin, 317 pp.; also Annual Sci. Rep., Contract NAS 5-11542, 1970–71, Space Science Engineering Center, University of Wisconsin, 166–270.

——, W. E. Shenk and W. Skillman, 1976: Wind estimates from cloud motions–Phase I of an *in situ* aircraft verification experiment. *J. Appl. Meteor.*, 15, 10–15.

——, —— and ——, 1977: Wind estimates from cloud motions: preliminary results of Phases I, II, and III of an *in situ* aircraft verification experiment. *J. Appl. Meteor.*, 16, 812–815.

Hubert, L. F., and L. F. Whitney, Jr., 1971: Wind estimation from geostationary satellite pictures. *Mon. Wea. Rev.*, 99, 665–672.

Kelly, N. O., and M. N. Zruber, 1973: Instrumentation aboard the Electra. *Atmos Tech.*, No. 1, 18–20.

Malkus, J. S., 1949: Effects of wind shear on some aspects of convection. *Trans. Amer. Geophys. Union*, 30, 19–25.

Minzner, R. A., W. E. Shenk, R. D. Teagle and J. Steranka, 1978: Stereographic cloud heights from imagery of SMS/GOES satellites. *Geophys. Res. Lett.*, 5, 21–24.

Peslen, C. A., 1979: Impact of short interval SMS digital data on wind vector determination for a severe local storms area. Submitted to *Mon. Wea. Rev.*

Rodgers, E., R. C. Gentry, W. E. Shenk and V. Oliver, 1979: The benefits of using short interval satellite images to derive winds for tropical cyclones. *Mon. Wea. Rev.*, 107, 577–584.

Smith, C. L., and A. F. Hasler, 1976: A comparison of low-cloud satellite wind estimates with analyses based on aircraft observations in a disturbed tropical region. *Mon. Wea. Rev.*, 104, 702–708.

Suchman, D., and D. Martin, 1976: Wind sets from SMS images: An assessment of quality for GATE. *J. Appl. Meteor.*, 15, 1265–1278.

——, —— and D. N. Sikdar, 1977: Deep convective mass transports: An estimate from a geostationary satellite. *Mon. Wea. Rev.*, 105, 943–955.

Telford, J. W., and P. B. Wagner, 1974. The measurement of horizontal air motion near clouds from aircraft. *J. Atmos. Sci.*, 31, 2066–2080.

Paper 43

Reprinted from JOURNAL OF THE ATMOSPHERIC SCIENCES, Vol. 36, No. 11, November 1979
American Meteorological Society
Printed in U. S. A.

Doppler Radar Study of a Warm Frontal Region

GERALD M. HEYMSFIELD[1]

Laboratory for Atmospheric Probing, Department of the Geophysical Sciences, The University of Chicago, Chicago, IL 60637

(Manuscript received 3 November 1978, in final form 2 July 1979)

ABSTRACT

This paper presents a case study of the structure of a warm frontal region as deduced from Doppler radar observations. The precipitation occurring ahead of the surface warm front was banded. The dominant precipitation bands were oriented transverse to the mid-level winds, and they were spaced ~110 km apart. It is suggested that these bands were formed by highly organized vertical circulations in a 2.5 km thick layer just above the warm frontal zone. The precipitation bands extended from this layer down to the surface. Near the surface additional circulations were produced by pressure perturbations resulting from cooling associated with melting snow. Some diagnostic calculations of ageostrophic winds, frontogenesis and vorticity production are presented. The frontogenesis calculation gives approximately a 2–4 h doubling time of the horizontal potential temperature gradient associated with the warm front, at mid-levels. The highly organized band-associated circulations suggest the importance of their inclusion in diagnostic calculations.

1. Introduction

The ability to forecast midlatitude weather is critically dependent on the ability to forecast the motion, formation, intensification and dissolution of fronts. In the literature, considerable attention has been given to a discussion of processes which maintain and intensify gradients of wind velocity and thermodynamic properties characteristic of frontal zones. Analytical and numerical studies have been directed to the question of frontogenesis—whether fronts can form in realistic times—and the physical mechanisms important for the frontogenesis (e.g., Hoskins and Bretherton, 1972; Mudrick, 1974; Gidel, 1978). Observations have also shed light on these questions through diagnostic studies of surface and elevated fronts (e.g., Reed and Sanders, 1953; Bosart, 1970).

Numerical simulations of the atmosphere typically use scales larger than about 100 km because of computer limitations. Temporal and spatial resolutions in synoptic studies of fronts are limited by the distribution and frequency of surface and upper air observations. Consequently, a basic question yet to be answered in the dynamics of fronts is the importance of small-scale motions (<100 km, say) on the large-scale dynamics. Is the effect so important that simulations of large-scale dynamics for extended periods of time cannot be successful unless small-scale processes are included in the conventional equations for large-scale dynamics? If so, we obviously need accurate parameterizations of the small-scale processes. In order to begin to answer these questions, we need observational descriptions of frontal regions at high space and time resolutions.

In recent years, methods of observations by Doppler radars (e.g., Miller and Strauch, 1974) have been developed which can provide the three-dimensional motion fields in regions of precipitation, down to scales of ~500 m. The limit on the scale at the large size end is imposed by the area that can be scanned by the Doppler radars—typically 50–100 km.

In this paper we shall present detailed observations of the three-dimensional wind and precipitation structure associated with a weak warm front which passed the Chicago area on 12 March 1976. The front was observed by a dual-Doppler radar system (consisting of the CHILL University of Chicago-Illinois State Water Survey radar and the CP-4 radar of the Field Observing Facility, National Center for Atmospheric Research) for a period of ~3 h as part of a winter storm experiment by the Laboratory for Atmospheric Probing, University of Chicago. The location of the radars is shown in Fig. 4a by the dots. The radar baseline is 51.35 km long and is oriented along 66–246°. Section 2 describes the synoptic conditions during the period of the radar observations. Associated with the warm front on 12 March 1976 were several precipitation bands. In Sections 3 and 4, we discuss the general features of these precipitation bands and their relation to the wind structure in the warm frontal region. In Section 5

[1] Present affiliation: NASA, Goddard Space Flight Center, Greenbelt, MD 20641.

0022-4928/79/112093-15$07.75

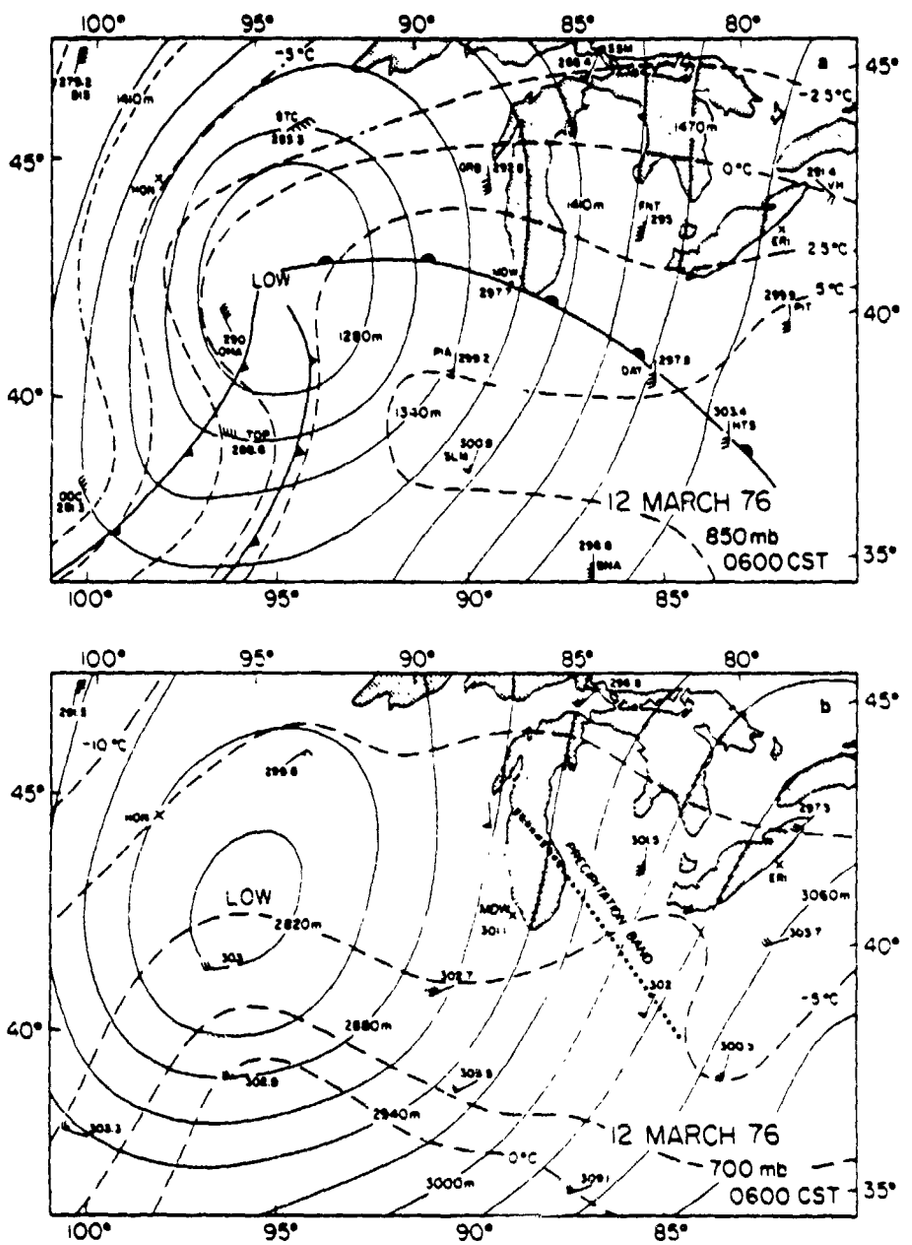


FIG. 1. Synoptic maps at 850 mb (a) and 700 mb (b). The height field in meters (solid lines) and isotherms in degrees Celsius (dashed lines) are shown on constant pressure surfaces. For the winds, a full barb is 5 m s^{-1} and a flag is 25 m s^{-1}. θ_E values are indicated near station locations, an × indicates missing wind or temperature data. The positions of fronts and stations are shown on the 850 mb level. These 0600 maps are actually at about 0525, i.e., 5–10 min after the sounding.

we present the detailed wind and reflectivity structure of one of the precipitation bands. Some tentative diagnostic calculations of ageostrophic winds, frontogenesis and vorticity production in the warm frontal region are given in Section 6.

2. Synoptic conditions on 12 March 1976

The radar observations presented in this paper were obtained from 0200 to 0500 CST (in the following, all times are CST) in a region of precipitation, the top of which was below 5 km (all heights are above ground level). The standard upper air observations were used to determine the synoptic-scale frontal structure at 0600; the release times for the soundings was about 0515, i.e., the end of the radar observations. A special low-level severe storm sounding was released at 0515 from Midway Airport, Chicago (MDW) which is located roughly

at the center of the radar analysis region. The special sounding does not measure winds, and above 700 mb, observations are taken only at 600 and 500 mb.

Fig. 1 shows the 850 and 700 mb maps for the central United States. The warm front had a northwest-southeast orientation in Illinois, with the surface low center ~500 km west of Chicago. Also shown is the equivalent potential temperature θ_E which is conserved under a pseudoadiabatic process, such as was probably occurring in the warm frontal lifting. On the surface map (which is not shown here, but presented later in Fig. 4a for a smaller region), the warm front was weakly evident from the surface temperatures, and more readily seen from the surface θ_E. This front moved northward at ~6 m s^{-1}, somewhat slower than the overall motion of the low center (~17 m s^{-1} from 245°).

Fig. 2 shows the distribution of temperatures and winds in the warm frontal zone in a vertical plane oriented perpendicular to the front. Soundings from Little Rock, Arkansas (LIT), Salem, Illinois (SLM), Peoria, Illinois (PIA), MDW, Green Bay, Wisconsin (GRB) and Sault-Ste.-Marie (SSM) were used in constructing this section. The low-level jet is quite strong with winds approaching 30 m s^{-1} above MDW. The warm frontal zone (Fig. 2a) has a slope of about 1/100 toward north, and is at a height of ~1.0 km above MDW. The resolution of the soundings was inadequate to depict the position of the eastern and leading edge of the mid-level warm air advection. Soundings from PIA, MDW and FNT (Flint, Michigan), shown in Fig. 3, indicate extremely dry air at PIA above 2 km height, which has not reached downstream to FNT. A moist, stable layer exists at ~3.5 km, with a temperature of about −11°C and 35% relative humidity. The MDW sounding shows a stable layer associated with the frontal zone (0.7–1.4 km), and between 2–3.6 km altitude. Potentially unstable layers occur at 2–3.5 km at PIA and 1.4–2 km at MDW, while potentially neutral layers are found above 3.7 and 6 km at these stations. The FNT sounding is both conditionally and potentially stable.

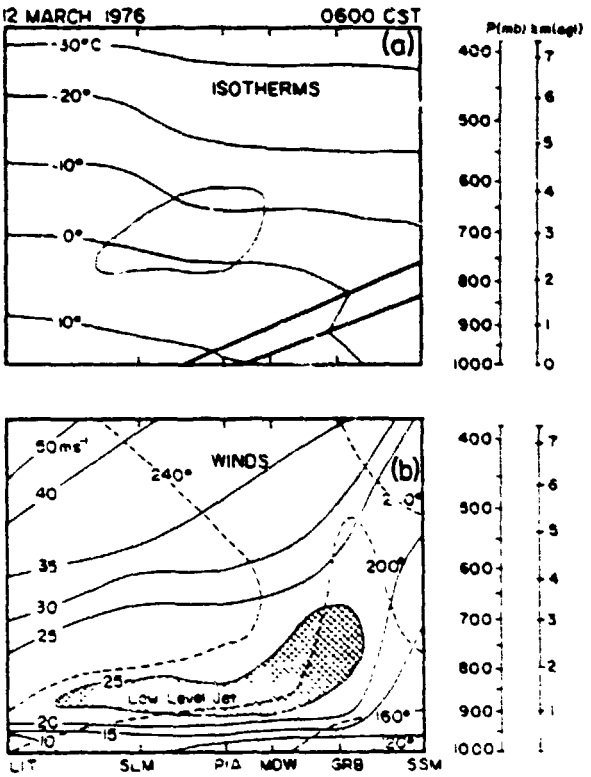


FIG. 2. Vertical cross sections of temperature (a) and winds (b) perpendicular to surface warm front. The stippling in (a) is a potentially unstable region; hatching in (b) shows winds ≥25 m s^{-1} associated with the low-level jet. Thick solid lines enclose warm frontal zone.

3. General features of the precipitation bands and cells

Near warm fronts, precipitation is commonly found in the form of bands having a wide range of dimensions, sometimes approaching hundreds of kilometers in length and 50 km in width (e.g., Browning and Harrold, 1969; Houze *et al.* 1976). The orientations of these bands are variable presumably due to differences in the mechanisms responsible for their formation.

The precipitation observed by the radars on 12 March 1976 was also predominantly banded in character with cores of high reflectivity embedded within the bands. The banded organization was obvious from an examination of the scope photographs of the CP-4, CHILL and conventional (Marseilles, Illinois) radar.

Fig. 4a shows the general area of precipitation near the surface at 0300. The heavy line encloses the region of precipitation detected by the Marseilles radar. The surface positions of the warm front at 0300 and 0600, the locations of the radars and the dual-Doppler analysis region (stippled) are also shown in this figure.

Fig. 4b shows the contoured radar reflectivity factor at 0° elevation angle at 0221. This figure was arrived at by an analysis of the CP-4 radar data. The 120 km by 120 km region was divided into a uniform grid with spacing 2 km. Reflectivities at the grid points were calculated by interpolation from the CP-4 observations. The region near the shores of Lake Michigan and the downtown Chicago area (indicated by vertical lines) was removed from the analysis to eliminate echoes from ground clutter. At 0221, the leading edge of the mesoscale precipitation region was just moving over the Lake. The axes of two bands of high reflectivity were identified in the figure. Band A1 had reflectivities in excess of 35 dBZ which, according to the empirical equation given by Gunn and Marshall (1958), corresponds to a rainfall rate of 3.0 mm h^{-1}. Band B1

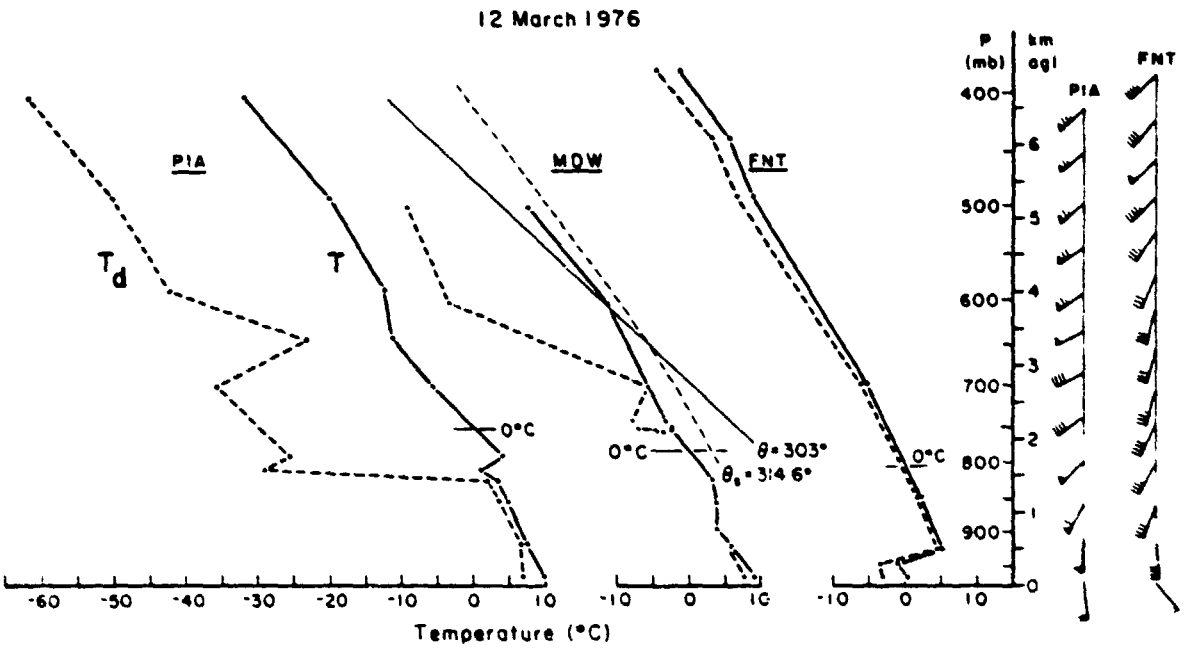


FIG. 3. Soundings from PIA, MDW and FNT. θ_e denotes pseudoadiabats. The release times of these soundings was about 0515.

has reflectivities in excess of 30 dBz, or a rainfall rate of ~7 mm h^{-1}.

None of the radars was able to see all the bands simultaneously or follow any particular band throughout its life time, because of radar range and sensitivity limitations. However, from a detailed examination of the radar scope photographs, reflectivity contour maps similar to the one shown in Fig. 4b, and other analyses, a broad picture of the distribution of precipitation was derived. Fig. 4c is a schematic summary of the bands as derived from compositing low-level reflectivity data from the three radars and adjusting them to a common time, 0300. This latter figure represents the bands at the initiation height of A bands, which for the present purpose is in the middle troposphere.

Two sets of bands A and B are identified on Fig. 4c. The first set (A1, A2 and A3) appeared as three bands embedded in a larger area of widespread precipitation. The orientation of these bands was 145–325° with a spacing of 100–110 km and their translational velocity was ~15 m s^{-1} from 240°. The second set of bands (B1, B2, etc.) had a typical spacing of about 40 km. The total number of these bands at 0300 was greater than the number of A bands, but their exact number is somewhat uncertain because of radar range and sensitivity limitations. The orientation of the B bands was 130–310°, and their translational velocity was ~22 m s^{-1} from 220°.

The first dominant band A1 moved across Chicago at about 0230. The radar observations showed the length of this band to be greater than 375 km. This band was not detected by the radars after 0330. However, a long band, believed to be band A1, was identified on the Defense Meteorological Satellite Program infrared picture taken at 0715. The time-adjusted position of this band has been shown on Fig. 1. Since band A1 was first observed at 2300 on 11 March 1976, it probably lasted for more than 8 h, strongly suggesting that it was associated with some quasi-steady feature of synoptic-scale dynamics.

Band A2 was weaker than A1 and passed through Chicago at about 0400. Band A3 was evident only as a line of discrete regions of high reflectivity on the 0° elevation Marseilles radar observations.

From the radar observations the B bands were deduced to have originated in generating cells at a height of ~5.2 km. Because of their low reflectivity, the generating cells were not detected in much of the radar observations taken at low elevation angles. However, they were seen in vertical incidence radar data and to a lesser extent in high-level CAPPI (constant altitude PPI) pictures of the kind shown in Fig. 4b. We suggest that there was a series of generating cells, aligned parallel to the orientation of the B bands, at a height of ~5.2 km. The precipitation trails emanating from these cells merged as the precipitation particles descended giving the appearance of bands in the low-level radar pictures. Calculations of particle trajectories [similar to those of Marshall (1953)] imply that the origin of the B bands was approximately 80 km upwind of their surface positions. The precipitation trail associated with a generating cell moves with the velocity of the generating cell (e.g., Marshall, 1953). The observed propagation velocity of the B bands is close to the

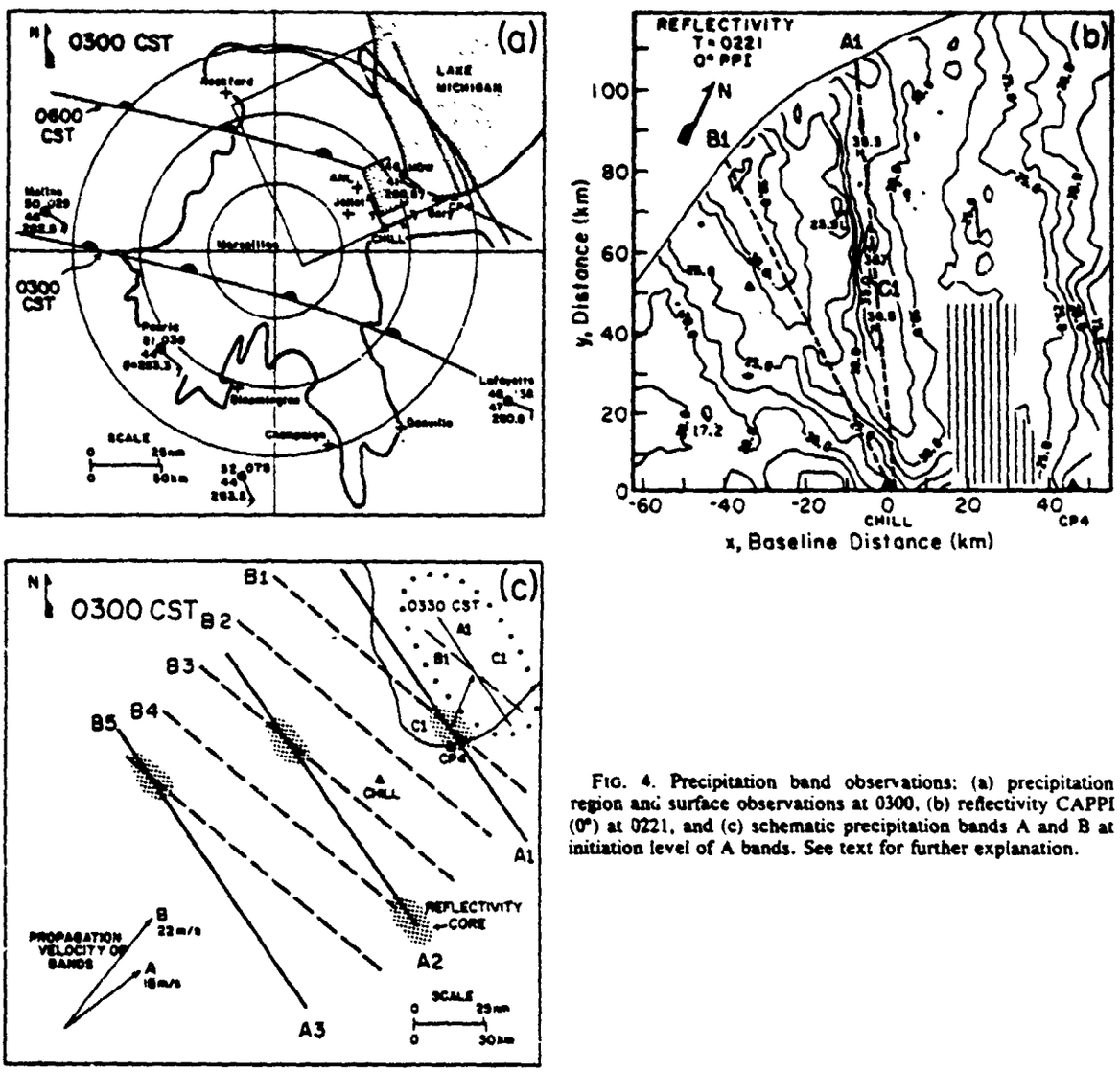


FIG. 4. Precipitation band observations: (a) precipitation region and surface observations at 0300, (b) reflectivity CAPPI (0°) at 0221, and (c) schematic precipitation bands A and B at initiation level of A bands. See text for further explanation.

wind velocity near the 5.2 km height (see Fig. 6b), supporting our contention that the B bands are "tied" to generating cells at a height of ~5.2 km.

As mentioned earlier a number of high-reflectivity cores were found to be embedded in band A. One such core C1 near the surface is indicated in Fig. 4b. These cores were found to move northward along band A. This behavior becomes explicable by the following consideration. High reflectivities are generated in regions where the precipitation trails associated with the B bands intersect the A bands at a height near the initiation level of the A bands; this is schematically depicted in Fig. 4c. The high reflectivity is probably due to the rapid precipitation growth produced by the injection of large particles formed by B bands aloft, into the A band. As the point of intersection of the precipitation trails and A bands moves, the region of high reflectivity moves northward along band A as illustrated by the inset in the top right corner of Fig. 4c. We do not see core C1 in Fig. 4b exactly at the intersection of A and B bands probably due to different particle trajectories associated with particles in the bands and reflectivity cores.

Further insight into the distribution of precipitation is provided by profiles of radar reflectivity factor and Doppler parameters at vertical incidence (Fig. 5) and the time-height cross section of contoured, average reflectivity factor (Fig. 6a). In arriving at Fig. 6a, the reflectivities observed by the CP-4 radar were averaged over the 27.5 km by 40 km

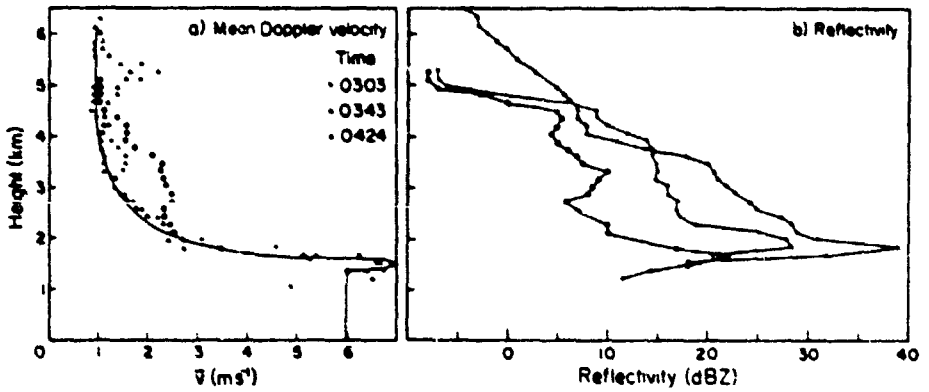


FIG. 5. Vertical incidence radar observations (collected from CHILL radar) at about 0303, 0343 and 0424. The points at each 150 m vertical interval represent an average over the 1–2 min interval of the data. Solid curve in (a) gives the V_T profile assumed in wind calculation: values at heights <1150 m were unreliable and therefore not plotted.

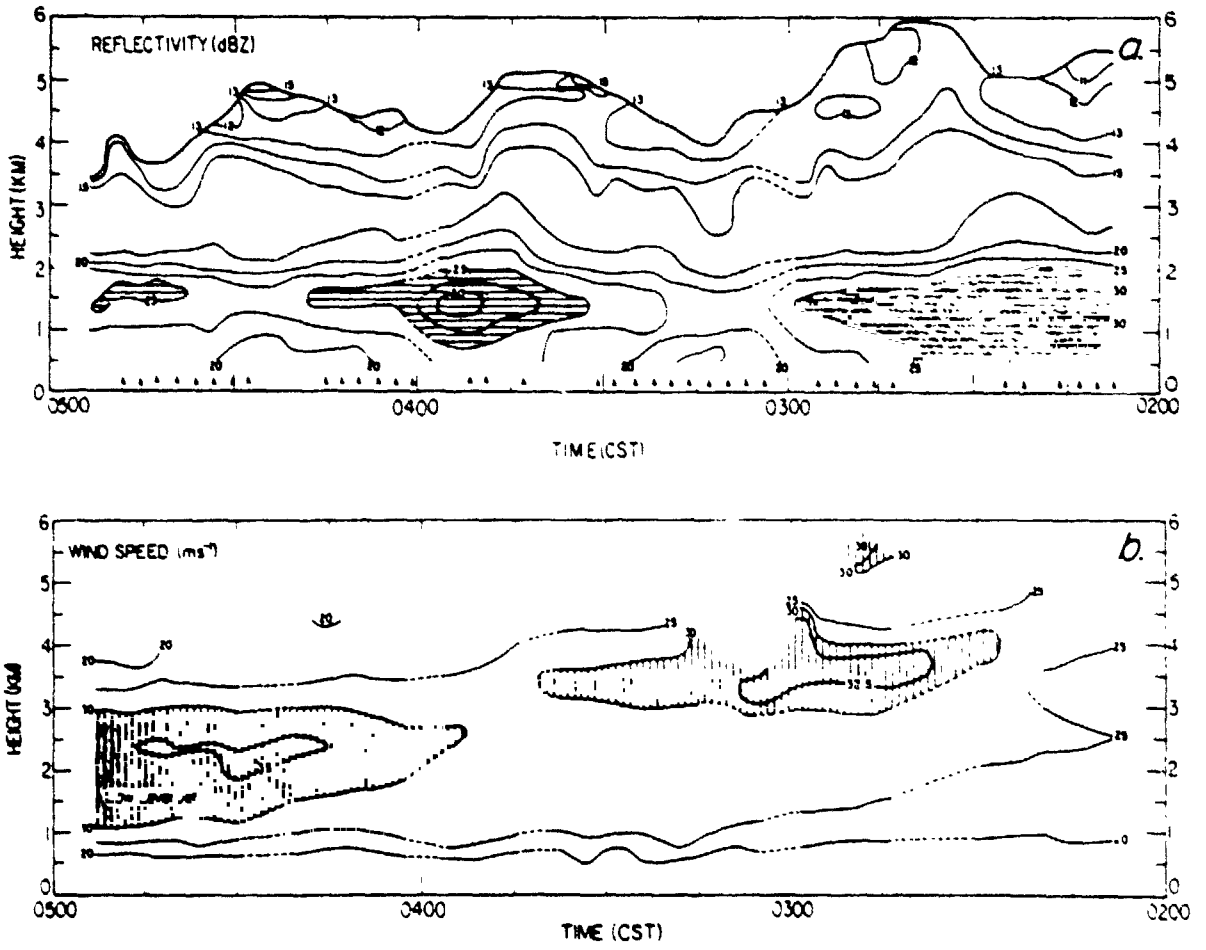


FIG. 6. Time height cross sections. Shown are (a) reflectivity, (b) wind speed, (c) wind direction, (d) vertical velocity (e) divergence, (f) vertical component of relative vorticity, (g) deformation and (h) vertical wind shear. Dashed contours indicate regions of missing observations. In (a) and (b), hatching indicates reflectivity ≥25 dBZ, and wind speed ≥30 m s^{-1}, respectively. Vertical hatchings in (d), (e), (f) give downdrafts ≤0 cm s^{-1}, and convergence and anticyclonic relative vorticity ≤ -2×10^{-4} s^{-1}, respectively. Horizontal hatching gives vertical motion ≥25 cm s^{-1}, and divergence and cyclonic relative vorticity ≥2×10^{-4} s^{-1}. (Cross-hatching in Fig. 6d shows updrafts ≥35 cm s^{-1}.) Deformation is displayed by a line whose orientation is parallel to the axis of dilatation (north toward top of figure), and length proportional to the rate of deformation. This cross section is centered over MDW.

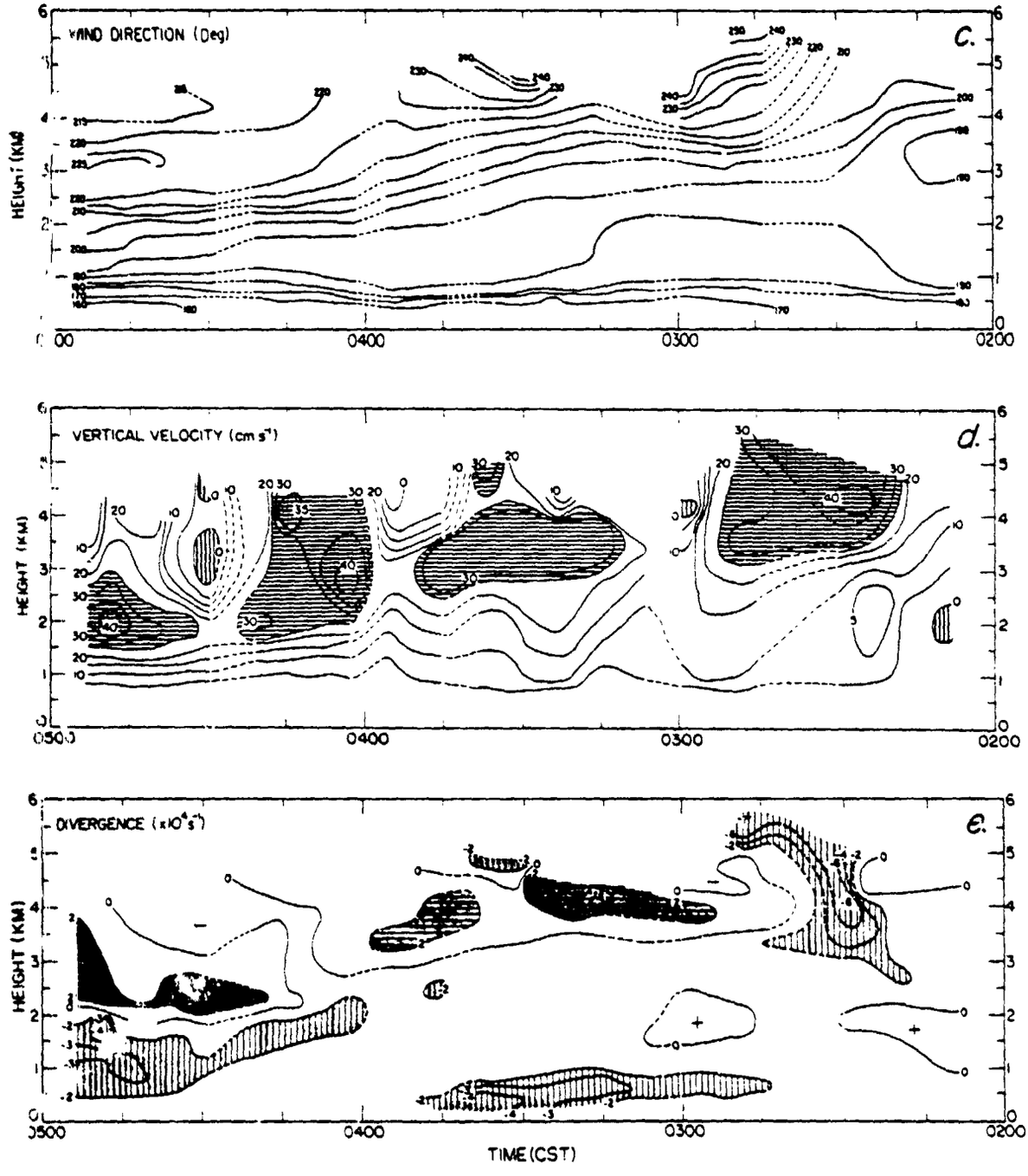


FIG. 6. (*Continued*)

dual-Doppler analysis region (shown stippled in Fig. 4a) over height increments of 100 m. A total of 45 radar volume scans (explained in the next section) was used to construct the time-height cross section of Fig. 6a; the locations of these volume scans are indicated by the arrows in Fig. 6a. Two pronounced mesoscale precipitation regions (identified by reflectivities ≥30 dBZ) A1 and A2 passed through the analysis region. The one centered at about 0230 (A1) is the dominant precipitation band mentioned earlier. The second precipitation band (A2) occurred at about 0345. A third mesoscale precipitation band was identified from the Marseilles radar observations as a set of discrete reflectivity cores. If

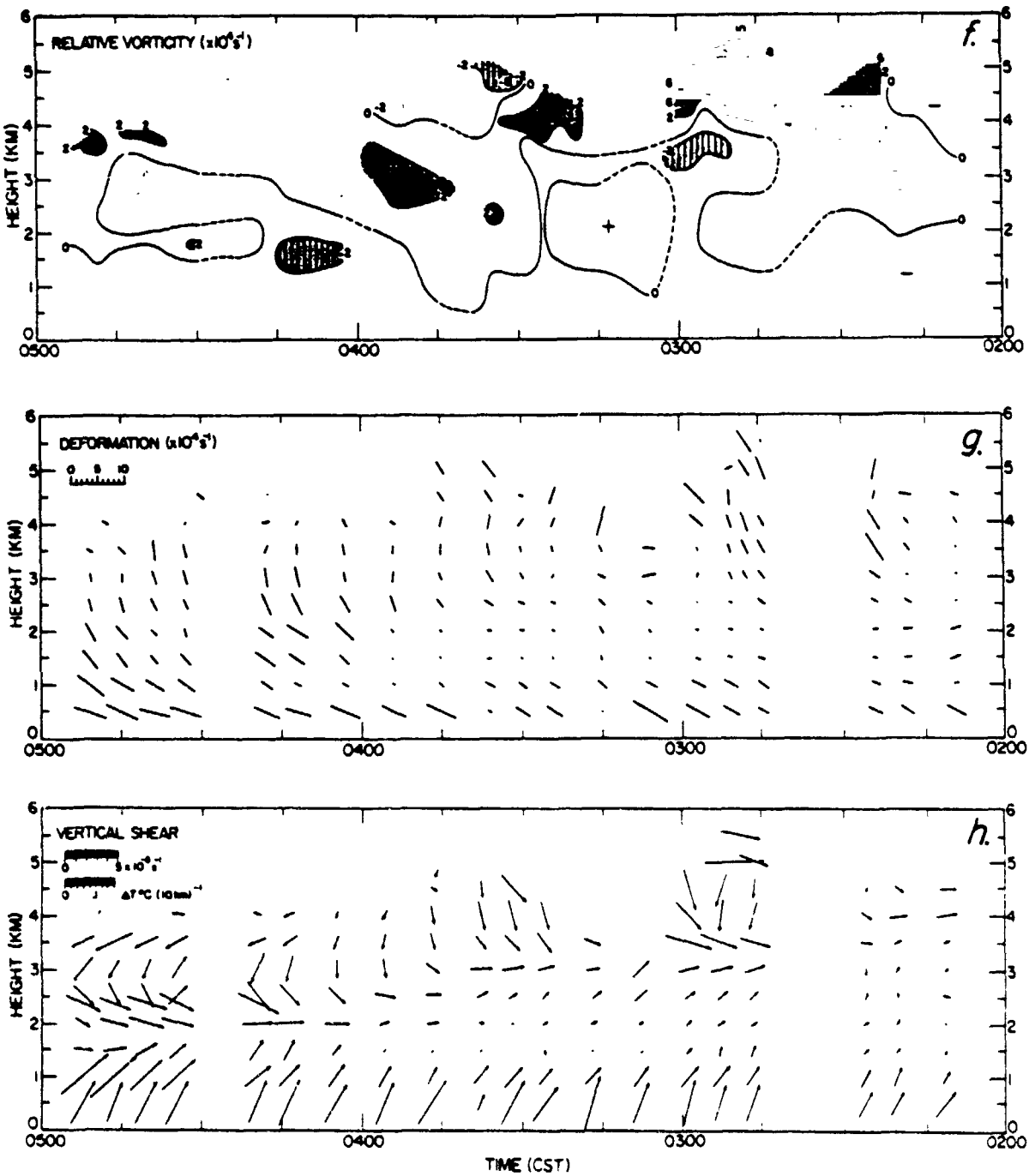


FIG. 6. (*Continued*)

represented on Fig. 6a, this precipitation region would occur beyond the left edge of the figure at about 0545. Other regions of high reflectivity are evident at a height of ~5 km. These are trails associated with generating cells giving rise to the B bands at lower altitudes. The melting layer is evidenced by the rapid increase in reflectivity below ~1.8 km height. This is in agreement with the temperature soundings (Fig. 3) which show the 0°C level at ~1.6–1.8 km above the surface.

The melting layer and the height variation of the reflectivity factor are more clearly seen in Fig. 5 which shows the mean Doppler velocity (positive downward) and reflectivity factor as obtained from

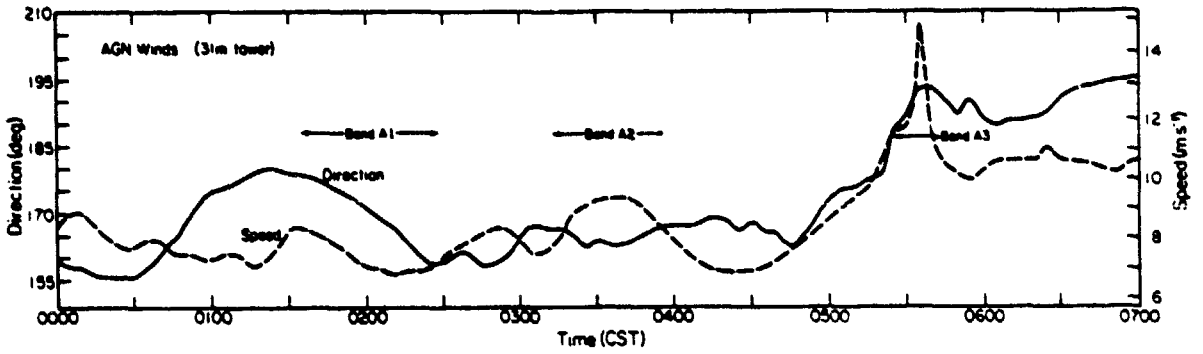


FIG. 7. AGN tower winds.

the vertical incidence observations by the CHILL radar at the three times shown. The melting layer is evidenced by the rapid increase in the radar reflectivity factor and the mean Doppler velocity near the 0°C level. It is interesting to note that the peak of the reflectivity factor occurred *above* the peak of the mean Doppler velocity $\bar{V}$ as would be anticipated from the theory of the melting band (e.g., Atlas, 1964). Also note the decrease of $\bar{V}$ below its peak, in association with the decreasing reflectivity factor; this is suggestive of particle breakup. Evaporation of particles is probably not occurring here since the low-level relative humidity is high (see Fig. 3). The vertical incidence observations detected precipitation to greater heights than shown in Fig. 6a because of their greater sensitivity. Also note the rather large values of $\bar{V}$ between heights of 2.5–3.5 km at 0424 and at ~5.2 km at 0303. These are suggestive of downdrafts of ~1–2 m s^{-1}, particularly near 5.2 km at 0303. Reference to Fig. 6a shows that this corresponds to the region between the two high-level trails. Downdrafts of such magnitudes were observed in association with generating cells by Heymsfield (1975) and Carbone and Bohne (1975).

4. General features of the wind distribution

a. Radar observations and analysis methods

In this section we shall present the general characteristics of the wind structure associated with the warm frontal region as deduced from the radar measurements. The radars scanned in the so-called coplane mode, that is, each radar scanned in a tilted plane whose rotation axis passed through the radar locations. The tilt of the coplane was varied in steps of 1° starting from 0° to an angle (~10°) large enough to top most of the precipitation region. A "volume scan," consisting of these coplanes, was completed in about 3 min. The radar scans were synchronized so that the two radars scanned the base and top coplanes approximately simultaneously. In addition to the coplane scans, some PPI and vertical incidence data were also collected.

The wind structure can be deduced from the Doppler radar observations on a much finer scale than is possible from conventional radiosonde observations. The principle underlying the derivation of the three-dimensional wind field (u,v,w) from the Doppler radar observations is very briefly as follows. Each Doppler radar measures the component of the precipitation velocity $(u, v, w - V_T)$ along the radial, where V_T is the average terminal fallspeed of the precipitation particles. Thus, we have four unknowns (u,v,w,V_T) and the two measured radial velocities. In order to derive u, v, w, two constraints are needed. In this work an assumption was made about V_T (see below) and the equation of continuity was used as the second constraint. Two techniques were used to derive the three-dimensional wind fields. A very brief description of these methods will now be given (details are given in Heymsfield, 1978).

In the first method, called the grid-point method, the analysis region centered at MDW was divided into a grid with horizontal spacings $\Delta x = \Delta y = 1.5$ km, parallel and perpendicular to the radar baseline, and vertical spacing $\Delta z = 0.7$ km. The radial velocities observed during a volume scan were assigned to a grid point using weighting and smoothing functions to filter scales of motion ≤3 km. The two radial velocities at each grid point were then combined to obtain u and v; here the assumption was made that V_T is given by the full curve in Fig. 5a and that the contribution of the vertical air velocity to the radial velocities may be neglected (radar elevation angles were usually less than 10°). Horizontal divergence fields were then computed and w was obtained by integrating the divergence vertically assuming $w = 0$ at the lowest observation level. Selected results of the grid point analysis are presented later (Figs. 10, 11).

In the second method, it was assumed that the wind field was linear over the analysis region, at any given height, i.e.,

$$\left.\begin{aligned} u &= u_0 + ax + by \\ v &= v_0 + cx + dy \end{aligned}\right\}, \qquad (1)$$

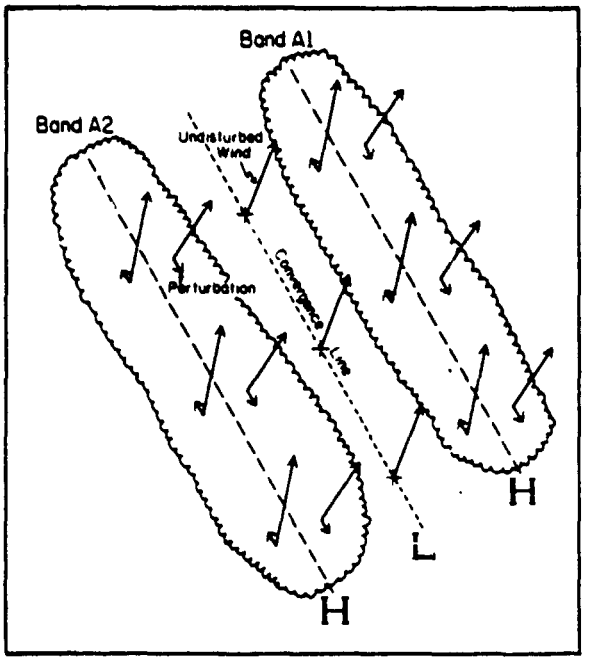


FIG. 8. Schematic of pressure perturbations and wind fluctuations below warm frontal zone. High (H) and low (L) pressure regions are shown.

where u_0, v_0, a, b, c and d are functions of the height. The analysis region was divided into 100 m thick layers, and the radial velocities from the two radars were used in a least-squares fitting procedure to determine u_0, v_0, a, b, c and d as functions of the height. From a, b, c and d, the horizontal divergence, vertical component of vorticity and deformation of the linear wind field were determined. The computed values represent means over the layer and, consequently, small-scale air motions ≤ 25 km have been filtered.

The horizontal velocities determined by the dual-Doppler method have very high accuracy for our purposes, where $|u,v| \leq 20$ cm s^{-1}; the vertical air velocities are not as accurate and $|w| \leq 20$ cm s^{-1} should be viewed with caution.

Before proceeding to a discussion of the wind distribution, we shall discuss the location of the warm frontal zone.

b. Location of the warm frontal zone

The location of the warm frontal zone is sketched in Fig. 9. The warm frontal passage was evident in the MDW surface observations taken at 0615, and 20–30 min earlier in 31 m tower measurements collected by the Argonne National Laboratory (AGN) which is located ~25 km southwest of MDW (see Fig. 7). As previously mentioned the frontal zone was at a height of 0.7–1.5 km above MDW at about 0520 (Fig. 3). These considerations give a frontal slope of $^1/_{100}$, similar to that shown in Fig. 2a.

c. Discussion of the wind distribution

A total of 38 dual-Doppler volume scans (indicated by the arrows in Fig. 6a) were used in the wind analyses. The results of the analysis, by the second method described in Section 4b, are presented in the time-height cross sections of Figs. 6b–6h which show the wind speed, wind direction, vertical air velocity, divergence of the horizontal wind, relative vorticity, deformation of the horizontal wind and the vertical shear of the horizontal wind, respectively. Dashed lines are used to indicate regions in which data from two or more volume scans were missing. Note that each calculated value represents an average over a 27.5 km by 40 km region. Taking account of the speed of motion of the system, this corresponds to a time span of ~30 min. Thus the gradients appearing in Fig. 6 are probably underestimates of the true gradients.

The sloping warm frontal zone is evident in Figs. 6b and 6c by wind speed in excess of 30 m s^{-1} and veering of the wind between 190° and 220°. The most striking feature of the warm advection above the frontal zone is the strong wind speed, with two distinct maxima ~5 m s^{-1} larger than the general frontal ascent. Note that these maxima occur between the three dominant precipitation bands A1, A2 and A3 (cf. with Fig. 6a; band A3 occurs beyond the left edge of this figure). In association with the fluctuation in the wind speed, there is a fluctuation in the wind direction, with a tendency for lower wind directions to occur above the A bands near the surface. Oscillations of smaller magnitude, both in the wind speed and direction, also occur below the melting band.

The vertical velocity in Fig. 6d shows a zone of upward velocity ~25–30 cm s^{-1} in a zone 1.5 km thick along the warm front. This implies a vertical displacement of about 3 km over a 3 h period, consistent with the position of the frontal zone in Fig. 9. Above the frontal zone, the vertical air velocity displays a distinctly cellular structure. Note that maxima of upward air motion occur slightly upwind of low-level reflectivity maxima associated with bands A1 and A2. Similarly, the downward air motions correspond to the minima of reflectivities between these bands. Here it should be mentioned that regions of downward air motions may not have been well observed because of the low reflectivities associated with them.

The maximum magnitudes of the divergence, deformation and vorticity are several times 10^{-4} s^{-1}, i.e., about an order of magnitude greater than typical synoptic-scale values. Particular attention is drawn to the occurrence of the two regions of large convergence in the lower levels between the three major bands A1, A2 and A3. This feature is discussed later.

The vorticity is generally positive, although there are some small pockets of negative vorticity (Fig.

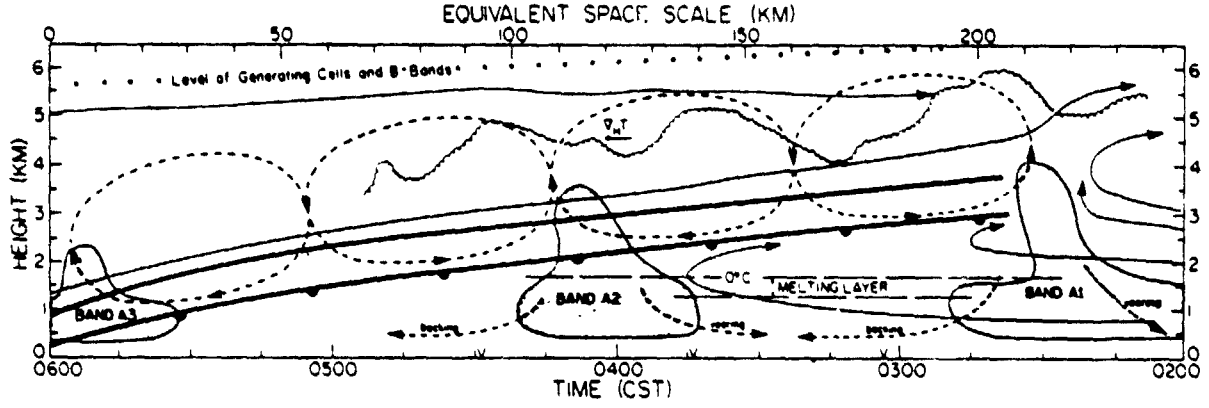


FIG. 9. Schematic of warm frontal structure over MDW. Thick solid lines enclose warm frontal zone. Streamlines are shown for large-scale front relative flow (solid lines), and secondary circulations relative to front (dashed lines). Radar cloud tops (scalloping) and time of vertical incidence observations (*V*) are shown.

6f). Note the transition in the direction of the deformation through the frontal zone.

The magnitude and direction of the vertical shear of the horizontal wind are shown in Fig. 6h. Again note the transition in the wind shear through the frontal zone. Above the frontal zone, we can apply the thermal wind equation

$$\partial \mathbf{V}_g/\partial z = \frac{g}{fT}\mathbf{k} \times \nabla_H T + \frac{1}{T}\frac{\partial T}{\partial z}\mathbf{V}_g \qquad (2)$$

to derive the horizontal gradient of the temperature $\nabla_H T$. In the above equation, $\mathbf{V}_g$ is the geostrophic wind velocity, z the height, $\mathbf{k}$ the unit vertical vector, g the acceleration of gravity and f the Coriolis parameter. We shall neglect the last term in the above equation. The resulting relationship between $|\partial \mathbf{V}_g/\partial z|$ and $|\nabla_H T|$ with $T = 273$ K is displayed in the top left-hand corner of Fig. 6h. Assuming $\partial \mathbf{V}_g/\partial z = \partial \mathbf{V}/\partial z$, where $\mathbf{V}$ is the measured horizontal wind velocity, we find $|\nabla_H T| \approx 0.1°\text{C}\,(10\text{ km})^{-1}$, with warmer air to the west-southwest (see $\nabla_H T$ in Fig. 9). The computed $\nabla_H T$ is in general agreement with the large-scale gradient of temperature in Fig. 1a (note the −5°C isotherm and the gradient of temperature implied by the temperatures across the isotherm in the general area of the radar observations). Also note that the shear vectors (and hence isotherms) and dilatation axes are generally parallel above the frontal zone. This implies a frontogenetic effect which will be discussed further in Section 6.

Some of the characteristic features of the reflectivity, wind and divergence patterns in the lower levels can be understood in terms of certain interactions between microphysics and dynamics first discussed by Atlas *et al.* (1969). The melting of ice particles below the 0°C isotherm cools the air and tends to produce a 0°C isothermal layer. Since the rate of cooling should increase with the rate of precipitation, we can anticipate greater cooling in association with the bands A1, A2 and A3. If the pressure above the 0°C level is undisturbed, higher pressures can be expected underneath the bands. This is schematically illustrated in Fig. 8. This distribution of pressure will give rise to perturbations in the air flow. We consider two extreme cases. In the absence of the Coriolis force, the region between the bands would be a region of convergence. The convergent flow would imply a backing and veering of the wind (relative to the mean horizontal wind) behind and ahead of the band, respectively. In the other extreme, we consider the perturbation winds to be completely adjusted to the Coriolis force. This is the case illustrated in Fig. 8. In this case, there would be no convergence but the winds would back and veer as in the previous case. In the real situation the perturbation winds would probably be in between the two extremes discussed considering the lifetime of the bands (many hours). Fig. 6e shows that there was pronounced convergence between the two bands. Backing and veering of the winds between the bands also occurred. This was verified by plotting the wind directions at a constant height below the 0°C level. A fluctuation of the anticipated kind was found superposed on a general trend of variation. The wind direction variation is also supported by the records of the AGN tower observations (see Fig. 7) which also shows the passage of the frontal surface near AGN.

From the above figures, we have constructed a schematic of the vertical circulations associated with the warm front (Fig. 9). The streamlines are perpendicular and relative to the A bands and were obtained from an analysis of Fig. 6. These show upgliding motion over the front. Secondary vertical circulations are shown above the warm front, as inferred from the maxima of wind speed and vertical motion in this region. Also shown are the circulations produced by melting of precipitation below the 0°C level.

It has been shown by Eliassen (1962) and Sawyer

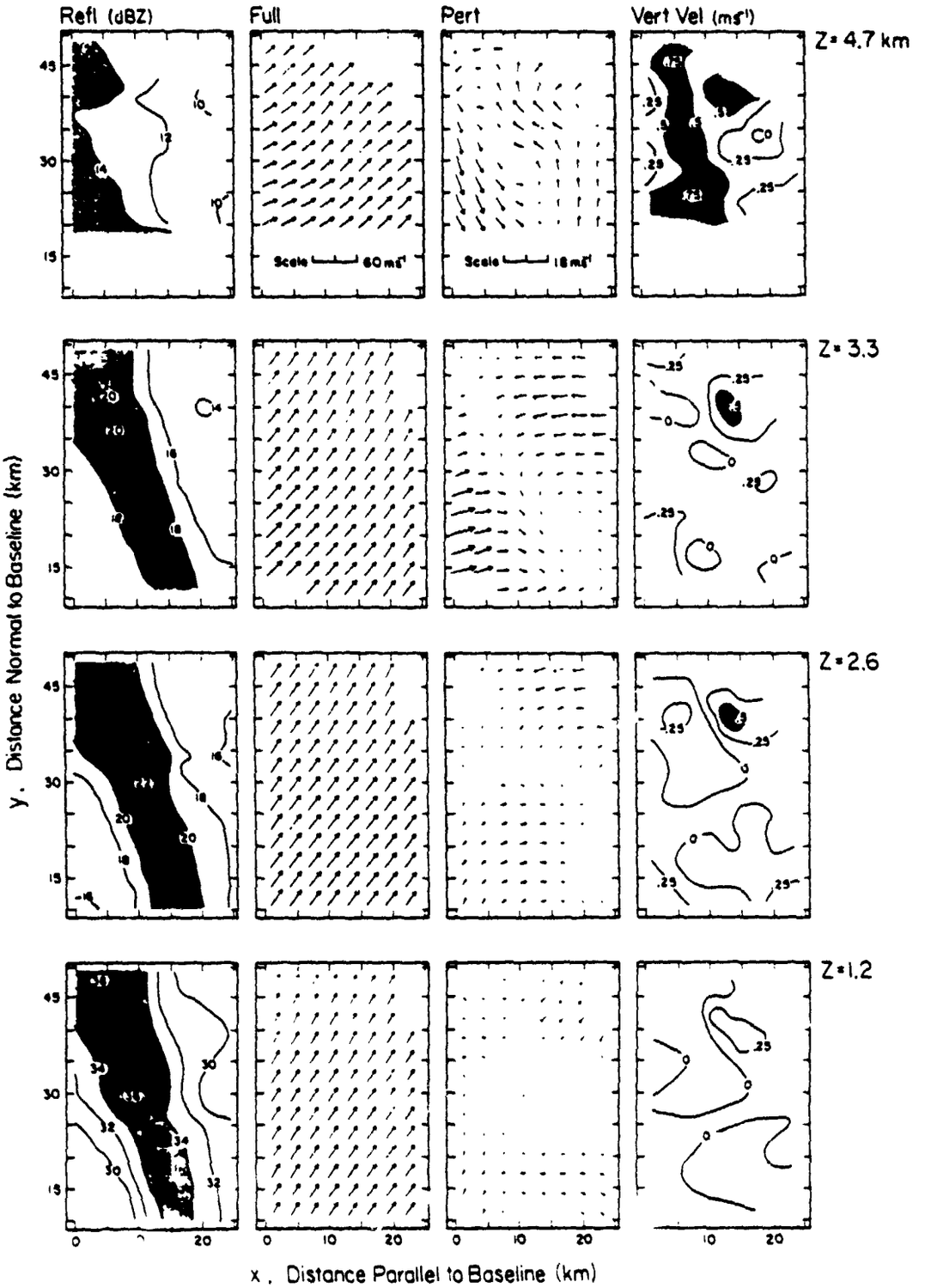


FIG. 10. Grid-point analysis at 0224 CST of three-dimensional winds and reflectivity. The analysis region is shown by small box in Fig. 4a. Stippling in reflectivity fields indicate band A1. Wind vectors are plotted at 3 km intervals; the perturbation wind vector was obtained by subtracting the layer mean wind vector. Vertical velocity ≥ 0.5 m s^{-1} is indicated by horizontal hatching. Center of analysis region is at MDW.

(1956) that in the presence of a deformation field acting to produce frontogenesis, vertical circulations must be set up, to balance the thermal wind. For warm fronts, Eliassen suggests that an indirect vertical circulation forms. The indirect vertical circulation associated with the warm front in the present case is perhaps similarly caused, although here we also have secondary vertical circulations with a 100 km horizontal scale superposed on the general warm frontal ascent.

5. Detailed structure of the precipitation bands

Fig. 10 shows the detailed structure associated with the precipitation band A1. The reflectivity and the winds are shown in this figure. Note the strong correlation between the band at 3.3 km, the vertical air velocity at 4.7 km, the line of convergence at 4.7 km, and the near discontinuity in the perturbation wind field at 4.7 km. Fig. 11 shows a summary at 0224 of the 3.3 km kinematic properties for band A1. Note the linear organization of the vorticity, deformation and horizontal shear along the band direction; the component of the band relative motion is directed along the band. Cyclonic vorticity is present downwind of the band. This figure shows the high degree of organized motion associated with the precipitation bands.

6. Calculations of ageostrophic winds, frontogenesis and vorticity production

Some tentative diagnostic calculations of the ageostrophic winds, frontogenesis and vorticity production will now be presented. As mentioned earlier, the horizontal winds have high accuracy, while the vertical velocity is less reliable. Also, thermodynamic measurements were limited to a much larger scale. Therefore, we are able to evaluate magnitudes of terms which consist of the horizontal winds, and with lesser accuracy the ones associated with the vertical wind component.

The ageostrophic wind is given by

$$\mathbf{V}_{ag} = \frac{1}{f}\hat{\mathbf{k}} \times \left(\frac{\partial \mathbf{V}}{\partial t} + u\frac{\partial \mathbf{V}}{\partial x} + v\frac{\partial \mathbf{V}}{\partial y} + w\frac{\partial \mathbf{V}}{\partial z} - \mathbf{T} \right), \quad (3)$$

where T is an acceleration due to turbulent momentum transfer. The radar observations were reanalyzed so as to compute time and space derivatives of the horizontal wind at 0224, 0254, 0318 and 0330 (details are given in Heymsfield, 1978) similar to the second method in Section 4a except using five volume scans. The ageostrophic wind was then calculated by assuming $\mathbf{T} = 0$; the geostrophic wind was computed from $\mathbf{V}_g = \mathbf{V} - \mathbf{V}_{ag}$.

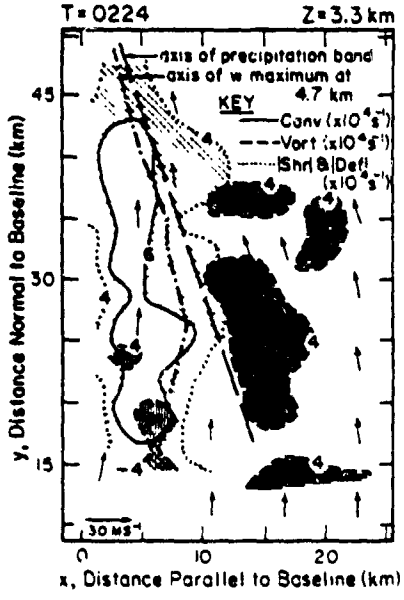


FIG. 11. Summary of 3.3 km altitude kinematic properties at 0224 for band A1. Winds relative to band A1 motion are superposed.

The values of mid-level ageostrophic and geostrophic wind were of the order of −50 and 70 m s^{-1}, respectively. The ageostrophic wind was such that the mid-level air flow upstream of the warm front (see Fig. 9) was strongly decelerated. The excessively high values of these winds are unrealistic, however. This points to the possibility of errors in the estimation of the derivatives in (3). Also, organized small-scale motions (such as those associated with A bands and smaller) represented by T, have been ignored in the calculation because of an inadequate knowledge of the turbulence structure. Rough estimates of T were made by considering vertical momentum transport by A bands. This term, however, was found to be of order 10^{-3} s^{-1}, and thus too small to balance other terms in (3).

During the period of the radar observations, the surface warm front in the 3 h maps was in an intensifying stage. After the radar observations, the 1800 synoptic maps showed an intensification of the warm front at 850 mb. The frontogenesis associated with the warm front was therefore examined according to the frontogenesis equation (Miller, 1948). This equation gives a diagnostic relation between the three-dimensional potential temperature gradient expressed by the frontogenesis factor F ($= d|\nabla\theta|/dt$), and the wind and potential temperature structure, where ∇ is the three-dimensional gradient operator. $F > 0$ implies frontogenesis. $F < 0$ corresponds to frontolysis. Since both the horizontal and vertical gradients of θ and the orientation of isentropes are inadequately known in the frontal zone, we attempt to show here some simple

deductions from Figs. 6e and 6g. The intensification of the magnitude of the horizontal gradient of θ from convergence and deformation is expressed by Haltiner and Martin [1957, Eqs. (17) and (18)] by

$$F = -\tfrac{1}{2}|\nabla_H\theta|(D\cos 2\beta + S), \qquad (4)$$

where S is stretching deformation, D is horizontal divergence and β is the angle between the axis of contraction and $\nabla_H\theta$. Frontogenesis is favored when there is convergence ($D < 0$) and isentropes are oriented within ~45° of the dilatation axis. Generally, $\nabla_H\theta$ is slightly larger than $\nabla_H T$ when the pressure surfaces are not highly sloped. Here, the assumption is made that $\nabla_H\theta \approx 0.1$ K (10 km)$^{-1}$, and of similar direction to $\nabla_H T$ in Fig. 9. Also, $\beta = 0$ is assumed as $\nabla_H\theta$ is probably perpendicular to the warm front and the axis of dilatation. From Figs. 6e and 6h, during the period 0212–0330 the frontogenesis contribution due to convergence and deformation was about 10–20 × 10^{-6} K s^{-1} (10 km)$^{-1}$ in the frontal zone region. Note that Fig. 11 suggests values approximately twice as large. By integration of (4), the doubling time of $|\nabla_H\theta|$ was estimated in the range 2–4 h. The contribution to F due to tilting of isentropes is also important on the scale of these observations. The tilting term is probably important in the observations because of the indirect circulation associated with the warm front. Very rough estimates from the soundings and Figs. 6d and 10 suggest that near the frontal zone, this term is approximately

$$-(\partial\theta/\partial z)(\partial w/\partial x)$$
$$\approx -5 \text{ K km}^{-1} \times -0.05 \text{ m s}^{-1}(10 \text{ km})^{-1}$$
$$= +250 \times 10^{-6} \text{ K s}^{-1}(10 \text{ km})^{-1}.$$

Thus, both terms are frontogenetic. The tilting term magnitude is about an order of magnitude larger than the convergence term, but should be viewed with caution because of the uncertainty in $\partial\theta/\partial z$ and $\partial w/\partial x$. Typical values reported by Bosart (1970) for mid-tropospheric frontogenesis in two dimensions were about 20 × 10^{-6} K s^{-1}(10 km)$^{-1}$, similar to the above estimates for a warm frontal case.

The vertical component of vorticity equation is given by

$$\frac{d(\zeta + f)}{dt} = -(\zeta + f)\nabla\cdot\mathbf{V} + \mathbf{k}\cdot\left(\frac{\partial\mathbf{V}}{\partial z}\times\nabla w\right)$$
$$-\mathbf{k}\cdot(\nabla\alpha\times\nabla p) - D. \qquad (5)$$

where $\alpha = 1/\rho$. The terms on the right side of (5) are the divergence, tilting, solenoidal and turbulence terms, respectively. Calculations of the divergence term using Figs. 6e and 6f give values in the range −4 to 11 × 10^{-6} s^{-2}, with largest positive values occurring within the frontal zone and particularly near the upward motion associated with band A1; the grid point method of Fig. 10 gave slightly larger values. Preliminary calculations of the tilting term using the grid point analysis (Fig. 10) gave magnitudes about +10^{-7} s^{-2} along band A1. The sum of the two terms deduced from Fig. 10 resulted in cyclonic vorticity production of order 10^{-7} s^{-2} slightly downwind of band A1. This magnitude of vorticity production acting on the same air parcel for 1 h would increase the absolute vorticity by ~3 × 10^{-3} s^{-1}. This is considerably larger than changes of vorticity suggested from Fig. 6. Note that we have neglected the solenoidal term—the vertical circulations and small-scale motions are probably important in regard to this term and also the tilting term. Also, the residence time of an air parcel in a strong vorticity production region may be shorter than 1 h, especially since the horizontal component of the upgliding air motion is ~10 m s^{-1} larger than the movement of the frontal zone.

7. Concluding remarks

The paper has presented dual-Doppler radar observations from a banded precipitation event associated with a warm front. The three major precipitation bands, having a spacing about 100 km (called A bands) were found to originate in a layer ~2.5 km thick located just above the warm frontal zone. Some insight regarding the relation between these precipitation bands and the dynamics of the warm front were deduced from the observations. Two secondary wind circulations were found to be associated with the A bands. The first type, located above the warm frontal zone, produced lines of vertical motion which presumably resulted in regions favorable to a more rapid growth of ice particles, which were descending from a higher level. Below the frontal zone and 0°C level, another circulation was apparently produced by pressure perturbations that were set up by the melting of ice particles, similar to that suggested by the theory of Atlas *et al.* (1969).

The mesoscale circulations deduced from the observations transport horizontal momentum vertically, and hence probably have an important effect on the large-scale dynamics. The diagnostic calculations of ageostrophic and geostrophic winds in Section 6 suggest that in further work, we should separate these mesoscale air motions from the large-scale flow, and carefully consider this mesoscale momentum transfer in the diagnostic calculations. The overall warm frontal circulation presented in Fig. 9 shows some resemblance to an indirect circulation for warm fronts, as suggested by Eliassen (1963), on which secondary circulations on a smaller scale are superposed. The long duration (~8 h) of the A bands suggests the deduced mesoscale circulations also have a long lifetime. Interestingly, these bands moved similar to motion of the synoptic

system, and ~15 m s^{-1} slower than winds in the initiation layer of A bands. The horizontal scale and phase speeds of these bands is similar to the observations of New England precipitation bands by Marks (1975). Lindzen and Tung (1976) explain Marks' observations in terms of ducted gravity waves. We were unable to verify whether the circulations here are due to gravity waves or an instability associated with the intensifying warm front because of inadequate thermodynamic data. These observations point to the necessity of highly detailed soundings perhaps every 15 min, and aircraft data, in conjunction with the radar observations.

Acknowledgments. The author is grateful to Professor Ramesh C. Srivastava for suggestions on the work and review of the manuscript. Dr. Alan Bohne is appreciated for helpful discussions. Messrs. Tim Lorello, Frank Hartney and John Landry assisted in the data reduction and drafting. Ms. Pam Travis is thanked for preparation of the manuscript. The author wishes to acknowledge the staff of the Illinois State Water Survey and the Field Observing Facility at NCAR, who were involved in the collection of the data. Argonne National Laboratory supplied tower observations used in the study. This research was supported by the Atmospheric Research Section, National Science Foundation, under Grants ATM-75-09683 and ATM-77-12616. Acknowledgment is made to the National Center for Atmospheric Research, which is sponsored by the National Science Foundation, for computer time used in this research.

REFERENCES

Atlas, D., 1964: Advances in radar meteorology. *Advances in Geophysics*, Vol. 10, Academic Press, 318–478.

——, R. Tatehira, R. C. Srivastava, W. Marker and R. E. Carbone, 1969: Precipitation-induced mesoscale wind perturbations in the melting layer. *Quart. J. Roy. Meteor. Soc.*, **95**, 544–560.

Bosart, L. F., 1970: Mid-tropospheric frontogenesis. *Quart. J. Meteor. Soc.*, **96**, 442–471.

Browning, K. A., and T. W. Harrold, 1969: Air motion and precipitation growth in a wave depression. *Quart. J. Meteor. Soc.*, **95**, 288–309.

Carbone, R. E., and A. R. Bohne, 1975: Cellular snow generation—a Doppler radar study. *J. Atmos. Sci.*, **32**, 1384–1394.

Eliassen, A., 1962: On the vertical circulation in frontal zones. *Geofys. Publ.*, **24**, 147–160.

Gidel, L. T., 1978: Simulation of the differences and similarities of warm and cold frontogenesis. *J. Geophys. Res.*, **83**, 915–928.

Gunn, K. L. S., and J. S. Marshall, 1958: The distribution with size of aggregate snowflakes. *J. Meteor.*, **15**, 452–466.

Haltiner, G. J., and F. L. Martin, 1957: *Dynamical and Physical Meteorology*. McGraw-Hill, 470 pp.

Heymsfield, A. J., 1975: Cirrus uncinus generating cells and the evolution of cirriform clouds. Part II: [illegible]cture and circulations of the cirrus uncinus gener[illegible]d. *J. Atmos. Sci.*, **32**, 809–819.

Heymsfield, J. M., 1978: A technique for ob[illegible] high vertical resolution mean wind and kinematic pro[illegible]m multiple Doppler radar observations. Tech. Rep. No. 42, Laboratory for Atmospheric Probing, The University of Chicago, 19 pp.

Hoskins, B. J., and F. P. Bretherton, 1972: Atmospheric frontogenesis models: Mathematical formulation and solution. *J. Atmos. Sci.*, **29**, 11–37.

Houze, R. A., P. V. Hobbs, K. Biswas and W. M. Davis, 1976: Mesoscale rain bands in extratropical cyclones. *Mon. Wea. Rev.*, **104**, 868–878.

Lindzen, R. S., and K. K. Tung, 1976: Banded convective activity and ducted gravity waves. *Mon. Wea. Rev.*, **104**, 1602–1617.

Marks, F. D., 1975: A study of the mesoscale precipitation patterns associated with the New England coastal front. M.Sc. thesis, Dept. of Meteorology, Massachusetts Institute of Technology, 42 pp.

Marshall, J. S., 1953: Precipitation trajectories and patterns. *J. Meteor.*, **10**, 25–29.

Miller, J. E., 1948: On the concept of frontogenesis. *J. Meteor.*, **5**, 169–171.

Miller, L. J., and R. G. Strauch, 1974: A dual-Doppler radar method for the determination of wind velocities within precipitating weather systems. *Remote Sens. Environ.*, **3**, 219–235.

Mudrick, S. E., 1974: A numerical study of frontogenesis. *J. Atmos. Sci.*, **31**, 869–892.

Reed, R. J., and F. Sanders, 1953: An investigation of the development of a mid-tropospheric frontal zone and its associated vorticity field. *J. Meteor.*, **10**, 338–349.

Sawyer, J. S., 1956: The vertical circulation at meteorological fronts and its relation to frontogenesis. *Proc. Roy. Soc. London*, **A234**, 346–362.

Paper 44

NASA Conference Publication 2076 – Fourth NASA Weather and Climate Program Science Review, January 24-25, 1979, NASA/GSFC, Greenbelt MD.

Paper No. 17

AN OBJECTIVE METHOD FOR FORECASTING TROPICAL CYCLONE INTENSITY AND MOTION USING NIMBUS-5 ESMR MEASUREMENTS AND NON-SATELLITE DERIVED DISCRIPTORS

H. E. Hunter, *ADAPT Services Corporation, Reading, Massachusetts*
E. B. Rodgers and W. E. Shenk, *Goddard Space Flight Center, Greenbelt, Maryland*

An empirical analysis program, based on finding an optimal representation of the data, has been applied to 120 observations of 29 1973 and 1974 North Pacific tropical cyclones. Each observation consists of a field of Nimbus-5 Electrically Scanning Microwave Radiometer radiation measurements at 267 grid points covering and surrounding the tropical cyclone plus nine other non-satellite derived discriptors. Forecast algorithms to estimate storm intensity and motion at 12, 24, 48, and 72 hours after each observation have been developed using an independent eigen screening analysis. These algorithms were based on best track data. Independent testing of these algorithms showed that the performance of most of these algorithms were better than persistence and the algorithms forecasting 24, 48, and 72 hour maximum wind speed were better than those made operationally by the Joint Typhoon Warning Center for 1973 and 1974 that did not use best track data.

Paper 45

NASA Technical Memorandum 80595 December 1979

AN OBJECTIVE METHOD FOR FORECASTING TROPICAL CYCLONE INTENSITY USING NIMBUS-5 ELECTRICALLY SCANNING MICROWAVE RADIOMETER MEASUREMENTS

Herbert E. Hunter

ADAPT Services Corporation

Reading, MASS 01867

Edward B. Rodgers and William E. Shenk

Goddard Space Flight Center

Greenbelt, MD 20771

ABSTRACT

An empirical analysis program, based on finding an optimal representation of the data, has been applied to 120 observations of 29 1973 and 1974 North Pacific tropical cyclones. Each observation consists of a field of Nimbus-5 Electrically Scanning Microwave Radiometer (ESMR-5) radiation measurements at 267 grid points covering and surrounding the tropical cyclone plus nine other non-satellite derived descriptors. Forecast algorithms to estimate the maximum wind speed at 12, 24, 48, and 72h after each observation were developed using three bases: the non-satellite derived descriptors, the ESMR-5 radiation measurements, and the combination of the two data bases. Independent testing of these algorithms showed that the average error made by algorithms developed from all three bases was less than the average error made by the persistence 24, 48, and 72h maximum wind speed forecast and less than the average errors made operationally by the Joint Typhoon Warning Center (JTWC) 48 and 72h maximum wind speed forecasts. The algorithms developed from the ESMR-5 base alone outperformed the JTWC operational forecast for the 48 and 72h maximum wind speed. Also, the ESMR-5 data base, when combined with the non-satellite base, produced algorithms that improved the 24 and 48h maximum wind speed forecast by as much as 10 percent and the 72h maximum wind forecast by approximately 16 percent as compared to the forecast obtained from the algorithms developed from the non-satellite data base alone.

Paper 46

Reprinted from *Geophysical Research Letters*, Vol. 5, No. 1, January 1978.

STEREOGRAPHIC CLOUD HEIGHTS FROM IMAGERY OF SMS/GOES SATELLITES

R. A. Minzner[1], W. E. Shenk[1], R. D. Teagle[2], and J. Steranka[3]

[1]Goddard Space Flight Center, Greenbelt, Maryland 20771
[2]Defense Mapping Agency, Topographic Center, 6500 Brooks Lane, Washington, D.C. 20315
[3]General Electric Company, 5050 Powder Mill Road, Beltsville, Maryland 20705

Abstract. Stereographic pairs of SMS/GOES images, generated simultaneously by the spin-scan cameras of each of two geostationary satellites (SMS 1 and SMS 2, separated by 32 degrees of longitude on February 17, 1975), have been analyzed photogrammetrically to yield cloud heights with a two-sigma uncertainty of 500 meters. The 32-degree angle between the image plane of the two satellites, plus the distortions involved in transferring the image of a nearly full hemisphere of the earth onto a plane, required the development of a special instrument to permit stereographic compilation. Cloud heights measured stereographically compared favorably with heights of the same clouds measured by radar and IR methods. The same SMS image pairs were used to measure mountain-top heights with a mean deviation of 0.24 km from cartographic values.

Background

The study of tropical cyclones and severe thunderstorms has been greatly enhanced by the availability of imagery of cloud fields observed from geosynchronous satellites such as ATS, SMS, and more recently GOES. The measurements of cloud-motion vectors from successive full-scan images obtained at the normal 30-minute interval by means of the ATS, SMS, and GOES sensors have permitted the inference of general wind fields over extensive portions of the globe [Young et al., 1972; Hubert and Whitney, 1974; Hussey and Shenk, 1976]. Limited-scan images obtained at a repetition rate equal to or less than 7.5 minutes from these satellites has led to more detailed circulation information [Gentry et al., 1976]. Accurate cloud-height information applied to the cloud-motion vectors would improve the circulation information, and could also benefit the determination of temperature and moisture profiles obtained from radiance sounding measurements. The simultaneous SMS/GOES infrared imagery which yields radiance temperatures of the tops of opaque clouds provides some information regarding the height of these cloud tops [Fritz and Winston, 1962; and Rao and Winston, 1963]. For a number of reasons the values of the cloud-top heights (particularly of cloud towers) inferred from IR data are subject to large uncertainties. These include the following: the limited spatial resolution of the IR sensor, 8 km compared with 0.8 km for the visible resolution [Hussey and Shenk, 1976]; the limited response time of the IR sensor [Negri, 1976]; the unknown opacity of various clouds [Rao, 1970; DeCotiis and Conlan, 1971; Shenk and Curran, 1973; and Young, 1975]; and the possible ambiguity in heights associated with temperatures obtained from temperature-height soundings, particularly for heights near or above the tropopause [Rao, 1970; and DeCotiis and Conlan, 1971]. Therefore, some other more precise methods for cloud-height determination would greatly enhance the utility of SMS/GOES imagery.

Paper number 7L0989.
0094-8276/78/017L-0989$02.00

Method and Results

The launch of SMS 2 in February, 1975 with an initial subsatellite point of about 107 degrees west longitude, while SMS 1 was located at 75 degrees west, provided an opportunity for generation of pairs of SMS images from which cloud heights might be extracted stereographically. (A subsatellite point of any particular satellite is the point of intersection of the surface of the earth and the line between the satellite and the earth's center). The suggestion of obtaining

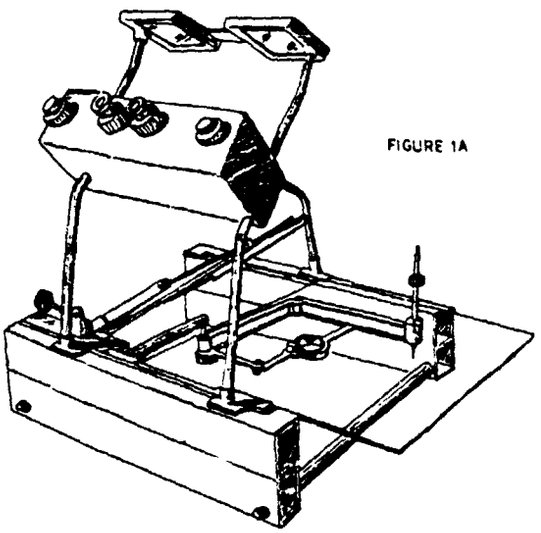


FIGURE 1A

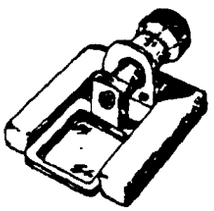


FIGURE 1B

Fig. 1A. Sketch of stereographic viewing and rectifying equipment consisting of a stereoscope mounted over a two-part viewing table, one part of which is biaxially tilcable.

B. Sketch of micrometer-coupled floating spot for measuring height in stereo model.

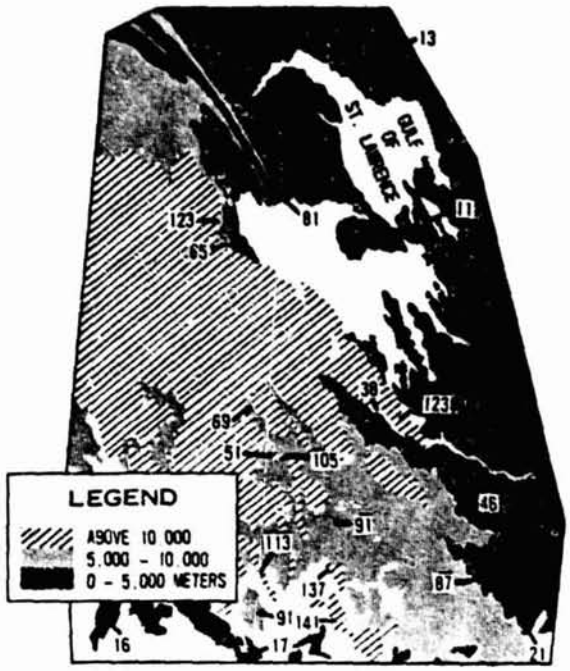


Fig. 2. Nova Scotia and surrounding areas including the stereographically measured heights (in hundreds of meters) of nearby clouds located as far as 67° of great-circle arc from the subsatellite point of the more remote satellite, SMS 2.

cloud heights from SMS images was fostered by the success at obtaining cloud heights from photographs obtained during one orbit of the unmanned Apollo-6 satellite [Shenk, 1971; Shenk and Holub, 1971; Shenk et al., 1975; Shull et al., 1976; and Shenk and Teagle, 1977].

Stereographic pairs of images were obtained by synchronizing the SMS-2 spin-scan camera with that of SMS 1 so that the northern-most limb of the earth was viewed by both satellites within a few seconds of each other. With this adjustment, both cameras continued to view the same latitude band of the earth within a few seconds of each other over the entire north to south stepping of the cameras. Thus, even though it takes 20 minutes for a full-disk image to be generated, the corresponding portions of the images generated by the two cameras represent essentially simultaneous pictures of identical swaths of cloud fields. This requirement for simultaneity is needed to overcome false cloud heights due to cloud motion, which would occur between non-simultaneous images.

The 32-degree angle between the image plane of the SMS-1 sensors and that of the SMS-2 sensors, plus the fact that both images represent a transformation of a near hemispheric surface onto a plane surface, caused problems in making the stereographical analysis. A specially designed mechanical-optical instrument depicted in Figure 1A provided a solution to these problems. This instrument consists of an Old Delft stereoscope mounted over a two-section light table, one section of which has a biaxial tilting mechanism, while the fixed section has a pantograph. The two images of the stereo pair are placed side by side in the proper sequence along the common x axis of both images. This axis is defined to be parallel to the line between the two subsatellite points. The image whose subsatellite point is nearest the cloud of interest is placed on the tilting table, which is tilted simultaneously in each of two mutually perpendicular axes, x and y, such that for some small area of interest, the y parallax is eliminated from the image pair. For this limited area, the model can be viewed stereoscopically by the operator, and true cloud heights can be deduced from the remaining x parallax. This is done by measuring the apparent height of the cloud in the stereo model with the micrometer-controlled reference spot depicted in Figure 1B, and scaling that model height to true height in accordance with an appropriate algorithm. This algorithm accounts for the curvature of the earth between the sea-level reference point, from which the cloud height is being measured, and the subsatellite point of the more remote satellite. With the aid of these devices, cloud-top heights were measured within a 2-sigma uncertainty of about ±500 meters, i.e., 90% of a series of repetitive readings of the same cloud feature agreed to within ±500 meters [Minzner et al., 1976, and Teagle, 1976].

The technique described above permitted the contouring of cloud fields over extensive portions of the western hemisphere. A sample compilation is shown in Figure 2. In this figure, depicting the Gulf of St. Lawrence and surrounding land masses including Nova Scotia, cloud

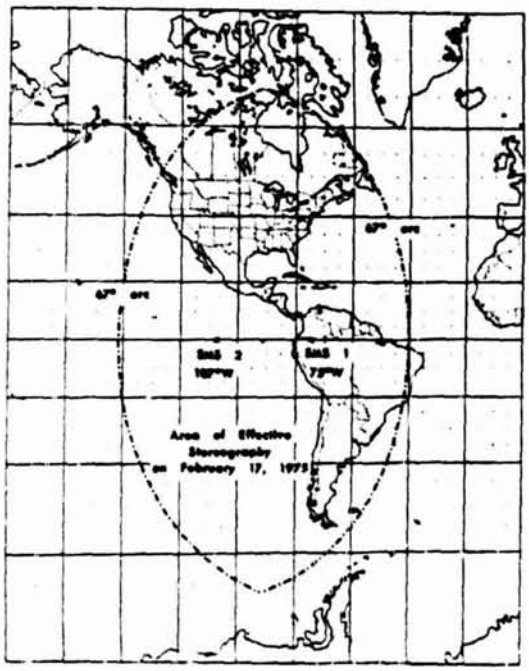


Fig. 3. Mercator projection of the Western Hemisphere showing limits of the area over which heights can be measured reliably from images generated simultaneously by SMS/GOES satellites over the specified locations.

masses are grouped into three height ranges, with the heights of particular features designated in hundreds of meters. The viewpoint of this image is from SMS 2 at 107 degrees west, and the compilation of cloud heights in this image provided the basis for estimating the limits of the region over which this particular technique of cloud-height stereography is applicable. This limit appears to be about 67 degrees of great-circle arc from the more remote subsatellite point, and permits the definition of the entire region over which the cloud-height stereography is applicable. For the location of the two satellites on February 17, 1975, this region is defined by the area on the earth's surface depicted in Figure 3.

Accuracy of Method

The stereographic analysis of an active storm over the southeastern United States provided an interesting set of cloud-top heights which in Table I are compared with a set of cloud heights obtained by ground-based radar and with a set of cloud heights obtained from satellite-measured infrared radiances. The general agreement and the nature of the mean systematic differences between these sets provides validity to the stereo heights. The radar heights are consistently found to be about one kilometer lower than the stereographic heights, with a mean difference of about -1.16 km, and a standard deviation of 0.83 km. This difference is accountable by the fact that radar echos of clouds represent primarily a reflection from precipitation rather than from the tops of visible clouds, and the upper portion of the cloud usually contains little or no precipitation. The infrared heights also are seen to be consistently lower than the stereographic heights with a mean difference of about -1.455 km and a standard deviation of 0.51 km. Again this difference is explainable by the physics of the infrared emissivity of clouds, whereby the infrared radiance is characteristic of some point within or below the cloud top rather than the top itself.

Further validity was given to these stereographic heights by the comparison of the stereographically measured values of the heights of 16 mountain peaks and one lake surface in the Peruvian Andes, with values published on topographic maps. These results are shown in Table II. These 17 comparisons show the stereo heights to be greater than the map values by a mean value of 0.24 km with a standard deviation of 0.36 km. These deviations, which suggest a small positive systematic error in the stereo values, represent errors in contour lines on the maps as well as the errors in the stereo measurements. The height of the surface of Lake Titicaca is of special interest since its stereo value was measured relative to a synthetic sea-level plane formed stereographically from two computer-generated grid systems which match the most precisely known shape of the earth.

A few cloud-height measurements made later with an improved version of the special mechanical-optical viewing instrument indicates that the two-sigma repeatability in cloud heights measured

TABLE I
Comparison of Radar- and IR-Derived Cloud-Top Heights with Stereographically Derived Values for Particular Clouds over Southeastern U.S. at 1900Z on February 17, 1975

Cloud	Stereo H (km)	Radar h (km)	Δh h-H (km)	IR h (km)	Δh h-H (km)
A	14.0	12.2	-1.8		
B	14.0	12.2	-1.8		
C	14.0	12.2	-1.8		
D	14.7	13.7	-1.0		
E	14.0	13.7	-0.3		
F	15.2	14.0	-1.2	13.0	-2.2
G	14.0	13.1	-0.9	13.0	-1.0
H	14.0	13.7	-0.3	12.25	-1.75
I	14.0	13.7	-0.3	13.0	-1.0
J	14.0	14.0	0.0	13.0	-1.0
K	14.0	13.7	-0.3	13.0	-1.0
L	13.0	----	----	12.2	-0.8
M	12.0	10.1	-1.9	10.1	-1.9
N	13.7	11.3	-2.4	----	----
O	13.0	10.7	-2.3	11.0	-2.0
P	12.0	----	----	10.6	-1.4
Q	12.0	----	----	10.1	-1.9

$\overline{\Delta h}$ = -1.164 km, $\overline{\Delta h}$ = -1.450 km,

σh = 0.833 km, σh = 0.509 km,

Δh = -1.16±0.83 km; Δh = -1.45±0.51 km.

TABLE II
Comparison of Stereographically Measured Heights of Mountains and a Lake in the Peruvian Andes with Values Cited on a Topographical Map

Point	Map Elevation h (km)	SMS Stereo Elevation H (km)	Δh H-h (km)
1	6.20	6.35	+0.15
2	6.40	6.10	-0.30
3	6.45	6.35	+0.10
4	6.63	6.70	+0.07
5	6.10	6.50	+0.40
6	5.80	5.95	+0.15
7	5.60	6.50	+0.90
8	5.90	6.40	+0.50
9	6.10	6.25	+0.15
10	6.61	6.50	-0.11
11	6.00	6.30	+0.30
12	5.80	5.68	-0.12
13	5.70	5.60	-0.11
14	5.90	6.10	+0.20
15	5.80	6.80	+1.00
16	5.80	6.40	+0.60
17*	3.81	4.11	+0.30

$\overline{\Delta h}$ = 0.236 km

σh = 0.360 km

Δh = 0.24±0.36 km

* Lake Titicaca

with optical-mechanical methods from SMS imagery may be as low as 350 m. This 25 percent improvement in the measuring accuracy results from two major instrument improvements: (1) a refined reference mark in the optical train replaces the original hand-held version; (2) a digital readout of the position of this reference mark eliminates the need to interpolate the x-displacement values. A plan for an improved determination of the absolute accuracy of the stereographic method by comparing stereographic cloud-top heights with cloud-top heights measured by dedicated aircraft has been prepared. This plan is awaiting implementation.

Conclusions

The work described above demonstrates that stereo imagery from two SMS or GOES satellites may be processed to provide cloud heights with sufficient accuracy for many meteorological purposes. The mechanical optical technique, however, requires the use of hard-copy prints or transparencies, and consequently is not suitable for a real-time or near-real-time operation. To circumvent this limitation, a man-interactive computerized method for extracting cloud heights from digital stereo imagery is being investigated.

References

DeCotiis, A.G., and E. Conlan, Cloud Information in three spatial dimensions using IR thermal imagery and vertical temperature profile data, Proceedings of the Seventh International Symposium on Remote Sensing of Environment, Willow Run Laboratories, University of Michigan, Ann Arbor, Michigan, 1971.

Fritz, S., and J.S. Winston, Synoptic use of radiation measurements from satellite TIROS II, Mon. Wea. Rev., 90, 1-9, 1962.

Gentry, R. C., E. Rodgers, W. E. Shenk, and V. Oliver, Deriving winds for hurricanes using short interval satellite imagery, Proceedings of Seventh Conference on Aerospace and Aeronautical Meteorology and Symposium on Remote Sensing from Satellites, Melbourne, Florida, November 16-19, 115-118, 1976.

Hubert, L. F. and L. F. Whitney, Jr., Compatibility of low-cloud vectors and rawins for synoptic scale analysis, NOAA Technical Report, NESS 70, p. 26, National Oceanographic and Atmospheric Administration, U.S. Department of Commerce, Washington, D.C., 19

Hussey, W. J. and W. E. Shenk, The GOES/SMS Users Guide, p. 118, National Environmental Satellite Service, NOAA, and Goddard Space Flight Center, NASA, Greenbelt, Maryland, 1976.

Minzner, R. A., W. E. Shenk, J. Steranka, and R. D. Teagle, Cloud heights determined stereographically from imagery recorded simultaneously by two synchronous meteorological satellites, SMS-1 and SMS-2, EOS, 57, 593, 1976.

Negri, A. J., D. W. Reynolds, and R. A. Maddox, Measurement of cumulonimbus clouds using quantitative satellite and radar data, Preprint Volume of Seventh Conference on Aerospace and Aeronautical Meteorology, Melbourne, Florida, 1976.

Rao, P. K. and J. S. Winston, An investigation of some synoptic capabilities of atmospheric "window" measurements from satellite TIROS II, J. Appl. Meteor., 2, 12-23, 1963.

Rao, P. K., Estimating cloud amount and height from satellite infrared radiation data, ESSA Technical Report, NESS-54, National Oceanographic and Atmospheric Administration, U.S. Department of Commerce, Washington, D.C., 1970.

Shenk, W. E., Cloud comparisons between Apollo-6 photography and ATS-3 and ESSA-3 photography, NASA TN D-6470, p. 47, Goddard Space Flight Center, Greenbelt, Maryland, 1971.

Shenk, W. E., and R. J. Holub, An example of detailed cloud contouring from Apollo-6 photography, Bull. Amer. Meteor. Soc., 52, 238, 1971.

Shenk, W. E., and R. J. Curran, A multi-spectral method for estimating cirrus cloud-top heights, J. Appl. Meteor., 12, 1213-1216, 1973.

Shenk, W. E., R. J. Holub, and R. A. Neff, stereographic cloud analysis from Apollo-6 photographs over a cold front, Bull. Amer. Meteor. Soc., 56, 4-16, 1975.

Shenk, W. E., and R. D. Teagle, Reply to comments by Holle and Ronne, Bull. Amer. Meteor. Soc., 58, 824-825, 1977.

Shull, C. W., R. D. Teagle, and R. G. Trip, Cloud elevations in near real time, Report of Defense Mapping Agency, Topographic Center to NASA, Goddard Space Flight Center, p. 121, GSFC, Greenbelt, Maryland, 1976.

Teagle, R. D., Cloud heights determined stereographically from SMS imagery, Report of Defense Mapping Agency, Topographic Center to NASA, Goddard Space Flight Center, p. 29, GSFC, Greenbelt, Maryland, 1976.

Young, M. T., R. C. Doolittle, and L. M. Mace, Operational procedures for estimating wind vectors from geostationary satellite data, NOAA Technical Memorandum, NESS 39, p. 19, National Environmental Satellite Service, National Oceanographic and Atmospheric Administration, U.S. Department of Commerce, Washington, D.C., 1972.

Young, M. T., The GOES wind operation, NOAA Technical Memorandum, NESS 64, 111-121, National Environmental Satellite Service, National Oceanographic and Atmospheric Administration, U.S. Department of Commerce, Washington, D.C., 1975.

(Received September 30, 1977; accepted October 25, 1977.)

Paper 47

NASA Conference Publication 2076 – Fourth NASA Weather and Climate Program Science Review, January 24-25, 1979, NASA/GSFC, Greenbelt, MD.

Paper No. 11

STEREOGRAPHIC CLOUD HEIGHTS FROM THE IMAGERY OF TWO SCAN-SYNCHRONIZED GEOSTATIONARY SATELLITES

R. A. Minzner, *Goddard Space Flight Center, Greenbelt, Maryland*
R. D. Teagle, *Defense Mapping Agency Topographic Center, Washington, D. C.*
J. Steranka, *General Electric Co., MATSCO, Beltsville, Maryland*
W. E. Shenk, *Goddard Space Flight Center, Greenbelt, Maryland*

Scan synchronization of the sensors of two SMS-GOES satellites yields imagery from which cloud heights can be derived stereographically with a theoretical two-sigma random uncertainty of ±0.25 km for pairs of satellites separated by 60° of longitude. Systematic height errors due to cloud motion can be kept below 100 m for all clouds with east-west components of speed below hurricane speed, provided the scan synchronization is within 4 seconds at the mid-point latitude, and the spin axis of each satellite is parallel to that of the earth.

Paper 48

NASA Technical Memorandum 79714 February 1979

IMPACT OF SHORT INTERVAL SMS DIGITAL DATA ON WIND VECTOR DETERMINATION FOR A SEVERE LOCAL STORMS AREA

Cynthia A. Peslen
Goddard Laboratory for Atmospheric Sciences
NASA/Goddard Space Flight Center
Greenbelt, MD 20771

ABSTRACT

The impact of 5 minute interval SMS-2 visible digital image data in analyzing severe local storms is examined using wind vectors derived from cloud tracking on time lapsed sequences of geosynchronous satellite images. The cloud tracking areas are located in the Central Plains, where on 6 May 1975, hail-producing thunderstorms occurred ahead of a well defined dry line.

Cloud tracking is performed on the Goddard Space Flight Center Atmospheric and Oceanographic Information Processing System (AOIPS). Lower tropospheric cumulus tracers are selected with the assistance of a cloud top height algorithm. Divergence is derived from the cloud motions using a modified Cressman (1959) objective analysis technique which is designed to organize irregularly spaced wind vectors into uniformly gridded wind fields.

The results demonstrate that satellite-derived wind vectors and their associated divergence fields complement conventional meteorological analyses in describing the conditions preceding severe local storm development. For this case, an apparent area of convergence consistently appeared ahead of the dry line and coincided with the developing area of severe weather. The magnitude of the maximum convergence varied between 10^{-5} sec^{-1} to 10^{-4} sec^{-1}. The number of satellite-derived wind vectors which were required to describe the kinematic properties of the low level atmosphere was adequate before numerous cumulonimbus cells formed. This technique is limited in areas of advanced convection.

NASA Conference Publication 2076 – Fourth NASA Weather and Climate Program Science Review, January 24-25, 1979, NASA/GSFC, Greenbelt, MD.

Paper No. 6

A COMPARISON BETWEEN GOES-1 IR DIGITAL DATA AND RADAR DATA FOR THE 4 APRIL 1977 SEVERE STORMS OUTBREAK

C. A. Peslen, *Laboratory of Atmospheric Sciences, Goddard Space Flight Center*
and Richard Anthony, *GE/MATSCO, Beltsville, Maryland*

ABSTRACT

The 4 April 1977 severe storms outbreak over Georgia and Alabama provided an excellent opportunity to examine the complementary characteristics between satellite and radar data.

Paper 50

Reprinted from *Monthly Weather Review*, 107, pp. 566-574, 1979.

The Computation of Isentropic Atmospheric Trajectories Using a "Discrete Model" Formulation

RALPH A. PETERSEN

Department of Geography, Northern Illinois University, DeKalb, IL 60115

LOUIS W. UCCELLINI[1]

Space Science and Engineering Center, University of Wisconsin, Madison, WI 53706

(Manuscript received 16 June 1978, in final form 29 January 1979)

ABSTRACT

An explicit technique for computing atmospheric trajectories, based on Greenspan's discrete model formulation, is presented as an alternative to the commonly used implicit scheme. The method provides an economical means of objectively obtaining computer-generated trajectories and accounts for the variable accelerations and local ψ tendencies along the entire trajectory path. The initial results presented show that the explicit computations are stable and very nearly energy-conservative. An application of the discrete model approach to a real data base and comparisons with trajectories determined by the implicit method yield favorable results, illustrating the utility of the explicit technique as a diagnostic tool.

1. Introduction

The construction of three-dimensional atmospheric trajectories provides a valuable diagnostic tool for illustrating and understanding three-dimensional transport and exchange processes associated with extratropical weather disturbances. The trajectory approach has been successfully applied to studies including the stratospheric and tropospheric air mass exchanges associated with a jet streak (Danielsen, 1968; Reiter, 1972), the dispersal of upper atmospheric pollutants and possible nuclear contamination (Reiter, 1972), the three-dimensional structure and moisture transport within extratropical cyclones (Danielsen, 1966, 1967; Buzzi and Rizzi, 1975), and the kinetic energy budget for an extratropical cyclone (Sechrist and Dutton, 1970). Trajectory models which use wind forecasts from operational primitive equation models at the National Meteorological Center (Reap, 1972) have also been developed to improve temperature and cloud forecasts.

The purpose of this paper is to present a new technique for constructing computer-generated trajectories as an alternative to the widely used implicit scheme developed by Danielsen (1961). Using this new approach, trajectories are computed by applying an explicit system of equations based on a "discrete model" theory developed by Greenspan (1972, 1973) to the atmospheric equations of motion. The method provides for objective, rapid and economical computations of trajectories, is stable for cyclonic and anticyclonic curvature, and explicitly accounts for the impact that spatial variations of the pressure gradient force have on parcel accelerations. The stability and accuracy of the method are tested by applying the explicit technique to an idealized case in which parcels undergo inertial oscillations. A diagnostic study utilizing standard rawinsonde upper air data is then presented to test the applicability of the explicit technique to a real atmospheric data base. Comparisons of selected trajectories computed by the explicit and implicit techniques are also shown.

2. The implicit technique for constructing atmospheric trajectories

Danielsen (1961) developed the most widely used method to compute atmospheric trajectories, for which he advocated the use of isentropic coordinates. Trajectories can be computed more accurately in the isentropic framework where two-dimensional, adiabatic flow implicitly includes vertical motions which would have to be calculated separately in Cartesian or isobaric coordinates. Trajectories are constructed on isentropic surfaces by si-

[1] Present affiliation: NASA/Goddard Space Flight Center, Laboratory for Atmospheric Sciences, Code 914, Greenbelt, MD 20771.

0027-0644/79/050566-09$06.25

multaneously solving energy and total displacement D formulas:

$$\frac{d}{dt}\left(\frac{U^2}{2}\right) + \frac{d}{dt}\psi = \frac{\partial}{\partial t}\psi, \tag{1}$$

$$D = \tfrac{1}{2}(U_1 + U_2)(t_2 - t_1), \tag{2}$$

in which U_1 and U_2 are the wind speeds at the initial and final times and ψ is the Montgomery streamfunction.

Although the implicit approach has been successfully applied to numerous diagnostic studies, the technique has several drawbacks. The simultaneous solution of (1) and (2) for anticyclonic trajectories does not always converge to a unique solution (Danielsen, 1961). Another difficulty involves estimating $\partial\psi/\partial t$ for the 12 h time interval normally imposed on trajectory calculations by the availability of upper atmospheric data. The large timestep and implicit nature of this method also make it difficult to account for variable parcel accelerations which occur along the trajectory path. Reiter (1972) attempted to include significant changes in the parcel accelerations by incorporating geostrophic departures within the iteration procedure. But Reiter's modification involves considerable subjectivity in estimating geostrophic departures at the beginning and end of the 12 h period that are subsequently applied along a significant portion of the trajectory path. In the following section, an alternative approach to diagnostic trajectory computations is presented which partially alleviates these problems.

3. An explicit method for computing atmospheric trajectories

The explicit method, based on the discrete model formulation of Greenspan (1972, 1973), calculates parcel acceleration, velocity and distance at successive time steps. The system of equations as applied to constructing adiabatic atmospheric trajectories is based on first determining the acceleration using the inviscid equations of motion

$$a_x^{(t)} = -\frac{\Delta\psi^{(t)}}{\Delta x} + f^{(t)}v^{(t)}, \tag{3}$$

$$a_y^{(t)} = -\frac{\Delta\psi^{(t)}}{\Delta y} - f^{(t)}u^{(t)}, \tag{4}$$

where a_x, u and a_y, v are the accelerations and velocity components in the x and y directions, respectively, f is the Coriolis parameter and superscripts refer to timestep number. The new velocity and distance components are then calculated by Greenspan's discrete formulation:

$$u^{(t+1)} = u^{(t)} + \Delta t[\tfrac{3}{2}a_x^{(t)} - \tfrac{1}{2}a_x^{(t-1)}], \tag{5}$$

$$v^{(t+1)} = v^{(t)} + \Delta t[\tfrac{3}{2}a_y^{(t)} - \tfrac{1}{2}a_y^{(t-1)}], \tag{6}$$

$$x^{(t+1)} = x^{(t)} + \tfrac{1}{2}\Delta t[u^{(t+1)} + u^{(t)}], \tag{7}$$

$$y^{(t+1)} = y^{(t)} + \tfrac{1}{2}\Delta t[v^{(t+1)} + v^{(t)}]. \tag{8}$$

A special "starter" formula is used for the first time increment

$$u^{(1)} = u^{(0)} + a_x^{(0)}\Delta t, \tag{9}$$

$$v^{(1)} = v^{(0)} + a_y^{(0)}\Delta t. \tag{10}$$

Readers may recognize the similarity of the $(\frac{3}{2}a - \frac{1}{2}a)$ term in (5) and (6) to the Adams-Bashforth scheme which was found to be quite stable (Lilly, 1965), but for small Δt only (Young, 1968). Greenspan (1972), however, found that in discretized applications the above explicit system conserves energy and is stable and accurate for a wide range of Δt.

To close the system of equations, new ψ values must be available at every timestep to calculate the spatial derivatives in (3) and (4). For diagnostic studies where ψ is available at 12 h intervals, $\psi^{(n)}$ is specified on a latitude-longitude grid by using the initial (0) and final (F) ψ values, and by assuming $(\partial\psi/\partial t)_{i,j}$ varies linearly at each grid point so that at any timestep n

$$\psi_{i,j}^{(n)} = \psi_{i,j}^{(0)} + \left[\frac{\psi^{(F)} - \psi^{(0)}}{t_F - t_0}\right]_{i,j} n\Delta t. \tag{11}$$

When ψ values are available 12 h before and after the time interval for which trajectories are constructed, a nonlinear approximation for $(\partial\psi/\partial t)_{i,j}$ is calculated using overlapping quadratic Lagrangian polynomials. This procedure, as presented by Bleck and Haagenson (1968) and expanded on by Whittaker and Petersen (1977), has the advantage that the derivatives of the resulting third-order polynomial are continuous and implicitly equivalent to second-order Taylor series approximations at the beginning and end of the period for which the trajectories are computed. Once $\psi_{i,j}^{(n)}$ is specified,

$$\frac{\Delta\psi^{(n)}}{\Delta x_{i,j}} \quad \text{and} \quad \frac{\Delta\psi^{(n)}}{\Delta y_{i,j}}$$

are calculated at grid points using standard centered differences over a $2\,\Delta x$ and $2\,\Delta y$ interval. The gradients are then interpolated from the grid-point location to the parcel position using the overlapping polynomial technique. With the gradients of ψ specified at the parcel position, the acceleration, new velocity and new position can be determined.

4. Initial experiments

Although the discrete model formulation conserves total energy for closed systems (Greenspan, 1972), the application of the technique to an open

system, in which the accelerations are approximated by the inviscid adiabatic equations of motion, and the interpolation of ψ gradients from grid points to parcel positions can introduce spurious energy sources. The initial experiments were designed to determine the magnitude of this spurious energy generation for a simplified, steady state flow. With steady-state assumptions, (1) reduces to

$$\frac{d}{dt}\left|\frac{U^2}{2}+\psi\right| = \frac{d}{dt}(T) = 0, \qquad (12)$$

so that the total energy (T) for a parcel remains constant. Any deviations from the initial total energy for a parcel represents a spurious energy source related to the discrete methodology and to the numerical interpolation technique needed for specifying ψ gradients at parcel positions.

For the initial experiments, a zonal ψ field and constant f (latitude of 40°N) were specified so that $dU_g/dt = 0$, where U_g is the geostrophic wind. Five parcels were chosen, with parcels 1 and 2 initialized with a subgeostrophic u component of 15.0 and 20.0 m s^{-1}, respectively. Parcel 3 was initialized with a geostrophic u component of 25.0 m s^{-1} and parcels 4 and 5 were initialized with supergeostrophic u components of 30.0 and 35.0 m s^{-1}. The initial v component was set equal to zero for all five parcels. With the ψ gradients kept constant in time, the parcel trajectories initialized with an ageostrophic wind should undergo an inertial oscillation (Newton, 1959) and therefore return to their respective original latitudinal position with a period equal to one-half pendulum day and with a wind speed equal to the original speed assigned to each parcel. The parcel initialized with a geostrophic wind speed should maintain a constant zonal component with v remaining equal to zero. The experiments were repeated with Δt set equal to 300, 600, 900, 1200 and 1800 s.

Fig. 1 illustrates the response of the five trajectories with Δt set at 900 s. The subgeostrophic parcels (1 and 2) initially turned to the left and accelerated, passed the equilibrium at which the u component is in geostrophic balance, and then began turning to the right while still accelerating. After 9 h, these parcels decelerated while continuing a turn to the right, as the v component became negative. The parcels then slowly turned to the left (after 15 h), completing the inertial oscillation by 18.75 h. The sequence of parcel accelerations was reversed for the supergeostrophic parcels (4 and 5). These initially ageostrophic parcels all displayed a period of oscillation equal to one-half pendulum day as expected. After one full oscillation and a return to within 50 m of their initial latitudinal position, parcels had velocities within 0.04 m s^{-1} of their original values. The trajectory for parcel 3, initialized with a geostrophic wind speed, remained zonal throughout the 24 h period.

Fig. 1 also illustrates that the supergeostrophic, subgeostrophic and geostrophic parcels were located at an equal distance from the starting position after the period of oscillation. The more supergeostrophic or subgeostrophic a parcel was initially the greater the respective subsequent decelerations or accelerations were in the x direction, thus neutralizing the different initial wind speeds. Applying the implicit technique of Eqs. (1) and (2) over the period of the inertial oscillation with an equal initial and final wind speed and without accounting for the ageostrophic departures yields a distance which varies for the different parcels, being greater for

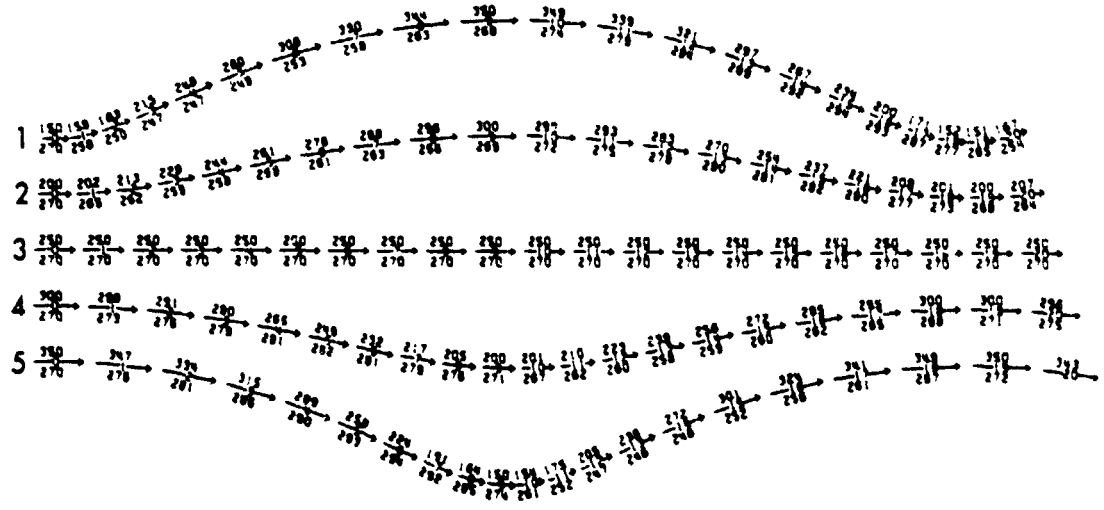


FIG. 1. Trajectories for various initial wind speeds with initial ψ field held constant. Trajectories 1 and 2 are initialized with a subgeostrophic wind, trajectory 3 is in geostrophic balance, and trajectories 4 and 5 are initialized with a supergeostrophic wind. The number above the arrow is wind speed, that on the arrow is the hour, and that below the arrow is wind direction. Trajectories are calculated with Δt = 900 s.

the supergeostrophic than for the subgeostrophic parcels. This example demonstrates the importance of objectively accounting for the geostrophic departures along the entire trajectory.

Energy diagnostics for parcels 1, 2, 4 and 5 are presented in Tables 1 and 2. The accumulated changes in total energy T over the first half of the inertial oscillation are given for various Δt (Table 1). The magnitudes of the accumulated changes in T were found to be a maximum at this time; whereas by (12), dT/dt should have been zero. In general, successive reductions in the Δt by 300 s increments for the trajectory computations reduced the magnitude of the spurious energy change by nearly half. But for all cases, the important factor is that the *maximum accumulated changes in T were two to three orders of magnitude less than the individual changes in the kinetic and potential energies for the same time period.* The accumulated change in total energy also reveals that, during the half-inertial period, subgeostrophic parcels experienced systematic gains in T while supergeostrophic parcels experienced systematic losses in T. The net changes in the total energy over the entire inertial period therefore approached zero.

The magnitudes of the changes in T during the first hour reveal the influence of the special "starting formulas" [(9) and (10)] that are needed for the first timestep (Table 2). The magnitudes of dT/dt during the first hour were of opposite sign and nearly two-thirds the magnitude of the accumulated change in T during the next 8 h period. The large errors are due to the nature of the starting formulas and to the initial conditions. With v initially set equal to zero, the u wind component is kept constant during the first time increment when it should be either increasing or decreasing as the ageostrophic parcels are subjected to latitudinal displacement. This error could be minimized by using a smaller initial timestep which would then be expanded in subsequent iterations until the standard time interval is achieved.

This initial experiment demonstrates the stability of the discrete formulation and the accuracy of the interpolation scheme in maintaining zonal geostrophic trajectories while also accounting for accelerations which result from initially unbalanced flow. Since the changes in the total energy which accumulated during successive iterations were at least two to three orders of magnitude less than changes in the kinetic and potential energies, the amount of error introduced to the trajectory paths by the discrete model approach should be minimal.

TABLE 1. Accumulated total energy changes (J kg^{-1}) from zero to 9 h for parcels 1, 2, 4 and 5 and Δt of 300 to 1800 s, with a geostrophic wind U_g of 25 m s^{-1}.

Δt (s)	Parcel's initial velocity 1 $0.6\,U_g$	2 $0.8\,U_g$	4 $1.2\,U_g$	5 $1.4\,U_g$
1800	+5.46	+2.28	−1.32	−1.74
1200	+2.35	+1.01	−0.57	−0.83
900	+1.32	+0.57	−0.30	−0.48
600	+0.62	+0.29	−0.10	−0.22
300	+0.24	+0.16	+0.07	−0.07

TABLE 2. As in Table 1 except from zero to 1 h.

Δt (s)	Parcel's initial velocity 1 $0.6\,U_g$	2 $0.8\,U_g$	4 $1.2\,U_g$	5 $1.4\,U_g$
1800	−2.10	−1.39	+2.15	+5.05
1200	−0.88	−0.60	+0.92	+2.18
900	−0.49	−0.33	+0.52	+1.21
600	−0.21	−0.14	+0.23	+0.53
300	−0.04	−0.03	+0.07	+0.13

5. Application of the explicit method to a diagnostic study

The explicit trajectory method was applied to an atmospheric data base to test the applicability of the discrete model formulation in diagnostic studies which rely on the rawinsonde network.

Nearly 100 trajectories were computed on the 330 K isentropic surface for the 12 h period 1200 GMT 10 May–0000 GMT 11 May 1973. The trajectories were initialized on a 2° latitude by 2° longitude grid within a domain bounded by 123°W 51°N, 83°W 51°N, 123°W 35°N and 83°W 35°N. The ψ gradients were specified utilizing the subjective ψ analyses shown in Fig. 2 and by assuming a 12 h linear ψ tendency at each grid point. The winds were initialized from subjective isotach and isogon analyses from 1200 GMT 10 May. The subjective isentropic analyses in this study were cross-checked with 15 vertical cross sections to incorporate the detailed vertical resolution of individual rawinsonde ascents within the horizontal wind analyses on isentropic surfaces (see Shapiro, 1970). An 1800 s timestep was used to calculate the trajectories.

At 1200 GMT 10 May 1973, three jet streaks on the 330 K surface (Fig. 2) were embedded within a general westerly flow stretching from the West Coast to the Great Lakes region. A deepening trough over the Great Lakes region yielded more cyclonic curvature to the flow field in the eastern third of the United States. By 0000 GMT 11 May, the two western jet streaks intensified slightly and propagated eastward while the eastern streak propagated northeastward and weakened (Fig. 2). The trough over the Great Lakes deepened slightly and was better defined by 0000 GMT.

Fig. 3 illustrates 20 trajectories which represent the larger sample of the trajectories computed. In general, the final wind speeds of the parcels were

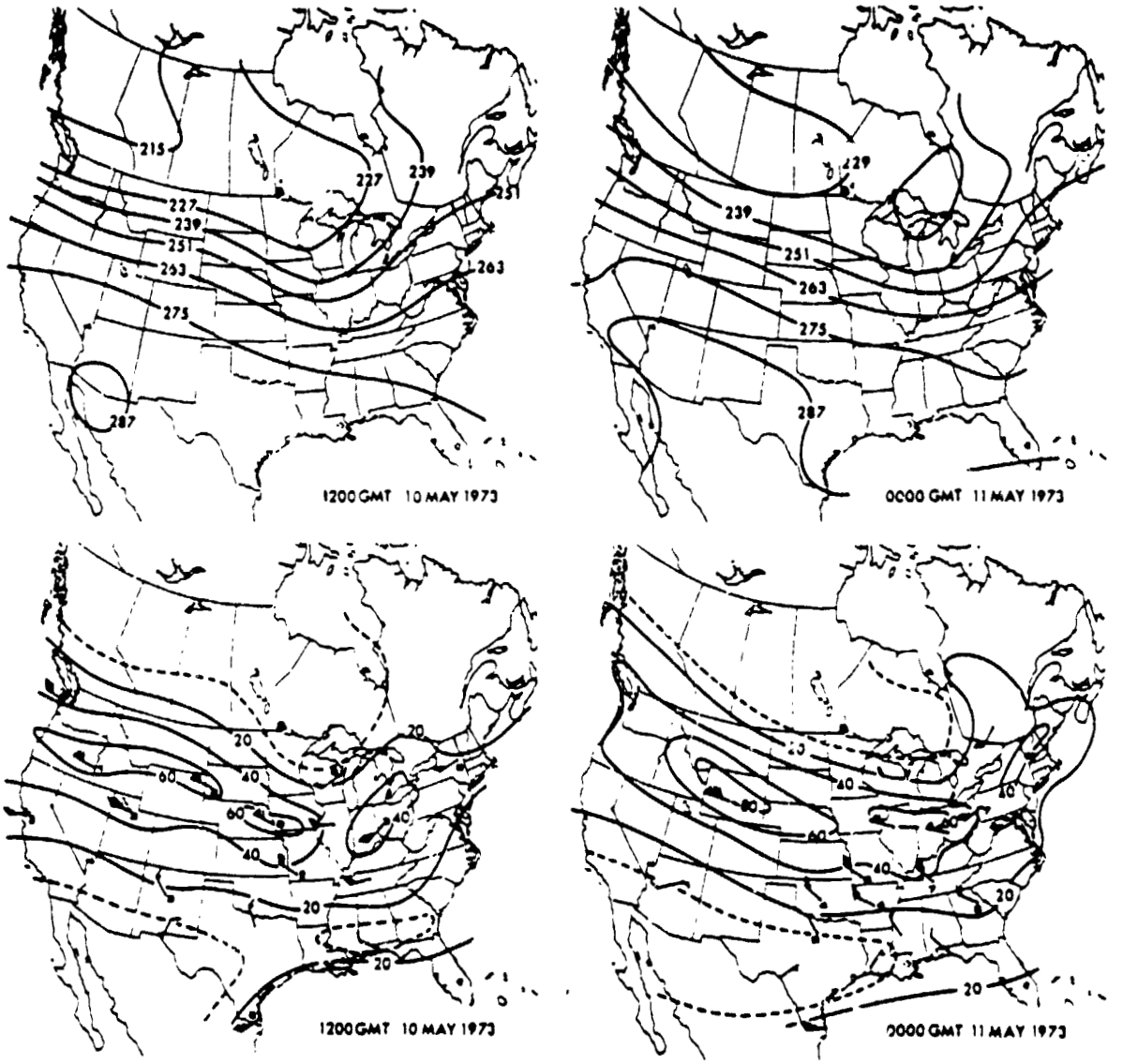


FIG. 2. 330 K isentropic analyses for 1200 GMT 10 May and 0000 GMT 11 May 1973. (top) ψ analysis ($275 = 3.275 \times 10^5$ m^2 s^{-2}). (bottom) isotach analysis (m s^{-1}). Wind barbs represent selected station wind reports rounded off to nearest 5 m s^{-1}, solid barb = 50 m s^{-1}, long barb = 10 m s^{-1}, short barb = 5 m s^{-1}. The number at the station is actual last digit of the wind report

within 5 m s^{-1} and 20° of the observed wind speed and direction, respectively. The final trajectory wind speeds were usually greater than the observed winds, which should be expected since frictional effects related to strong horizontal and vertical wind shears are not included in the scheme.[2]

The trajectories initialized within the jet streak over the western United States moved rapidly eastward while accelerating slightly, then turned to the right and began decelerating by 0000 GMT 11 May. Trajectories E and P, initially located to the south of the western jet streak, accelerated toward the region of strong ψ gradients and maximum winds. On approaching the deepening trough over the Great Lakes region, trajectories G and H turned cyclonically and accelerated in response to the increasing ψ gradient.

Trajectories L and Q are included to illustrate the ability of the explicit technique to compute trajectories within weaker and more variable flow regimes which could result in more erratic parcel

[2] Departures are often observed for trajectories originating very close together in regions of strong winds and large horizontal and vertical wind shears. In those regions, the absence of a direct tracer in the atmosphere (radioactive debris, etc.) and questions regarding the conservation of potential vorticity (Shapiro, 1978) make it difficult to ascertain the representativeness of parcel trajectories computed by any scheme. Also, when large-scale upward motions are present, the parcels may undergo extensive non-adiabatic lifting, which further complicates comparisons.

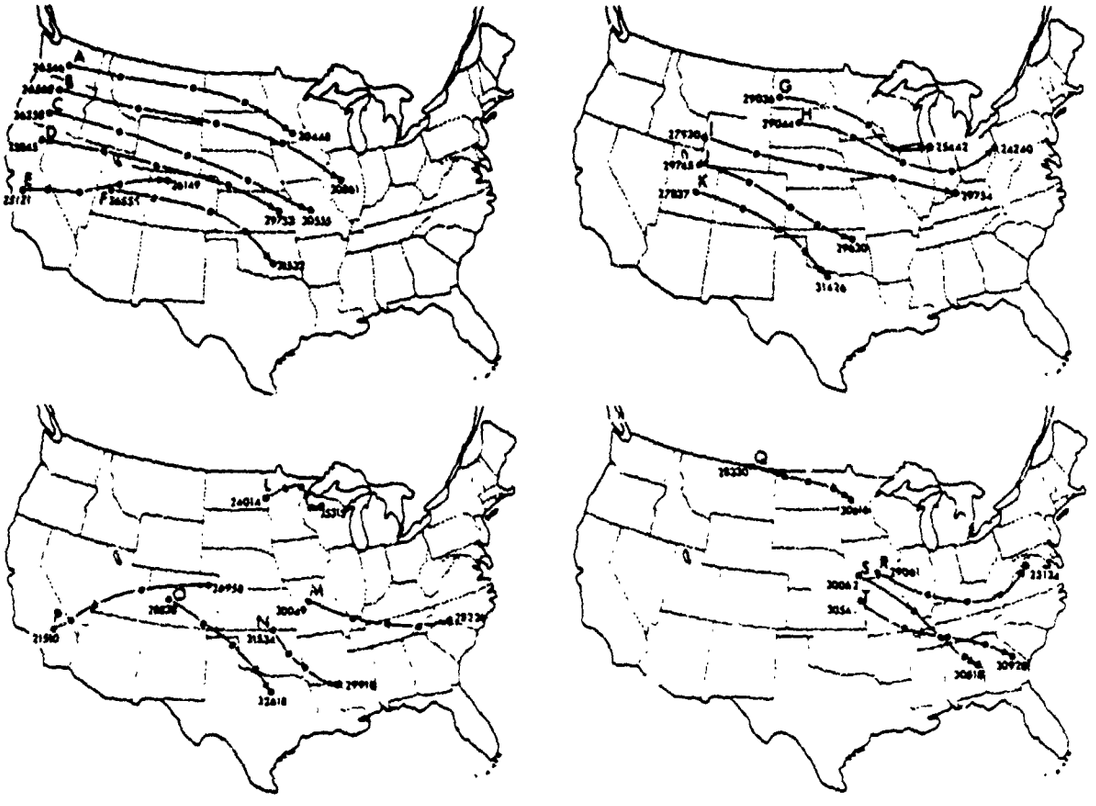


FIG. 3. Trajectories computed for 1200 GMT 10 May through 0000 GMT 11 May 1973 on 330 K isentropic surface utilizing discrete model technique. Positions at 3 h intervals are indicated by ⊕. The initial and final wind direction and speed are indicated (31426 = 314° 26 m s^{-1}).

movement. Trajectories M, N, O and R, S, T depict the parcel responses upon exiting the jet streaks originally extending into the Great Plains (Fig. 2). These parcels turned to the right and decelerated upon exiting the streak. Several of the parcels (R and M) then turned to the left and accelerated later in the 12 h period upon approaching the trough line which extended from the Great Lakes to the southeastern United States. Trajectories R and S depict how critical initial wind speed and the magnitude of the ageostrophic component were to the trajectories originating in the exit region of the jet streak. These same trajectories were originally only 222 km apart, yet terminated nearly 1000 km apart 12 h later. Both trajectories appear to be reasonable in that the final wind speeds were within 2 m s^{-1} of the observed wind speeds and 10° of the observed directions at 0000 GMT 11 May.

Fig. 4 illustrates selected comparisons made between trajectories computed with the explicit discrete model formulation, for which the final computed velocities agreed closely with 0000 GMT observations, and Danielsen's implicit scheme [(1) and (2)]. In general, the discrepancies were minor as the difference in the 12 h positions determined by the two methods fell well within the error range expected from data acquisition and analysis errors (Reiter, 1972). However, the two schemes did yield significantly different paths for the parcels (R and S) originating in the exit region of the jet streak (Fig. 4C). In the first case, the explicit trajectory (R, dashed) originated in southeast Iowa, decelerated and turned to the right upon exiting the streak and later accelerated, ending near Washington, DC, with a wind speed and direction nearly equal to the wind observation near that location (Fig. 2). The implicit scheme did not incorporate these variable accelerations along the path of the trajectory since only the average wind speed based on initial and final values was used to determine the distance traveled (2). The implicit scheme therefore yielded an energy-consistent trajectory which extended 330 km beyond the endpoint of the trajectory computed with the explicit technique. In the second case, the trajectories originating in southwest Iowa (S) show even more of a discrepancy. The explicit trajectory decelerated rapidly from an initial 64 m s^{-1} wind speed and turned to the right, ending in Alabama. The

FIG. 4. Trajectories calculated using Danielsen's implicit scheme (solid) and using discrete model formulation (dashed) for period from 1200 GMT 10 May through 0000 GMT 11 May 1973. Final wind direction and speed are indicated for trajectories computed using discrete model formulation (27430 = 274° 30 m s^{-1}).

final wind speed of 18 m s^{-1} and wind direction of 305° were within 2 m s^{-1} and 10°, respectively, of the observed wind near that location. Our application of the implicit scheme did not give this trajectory but yielded an energy-consistent trajectory ending near Cape Hatteras, North Carolina, 1000 km from the termination of the explicitly computed trajectory.

While the final wind speeds were approximately equal for both the explicit and implicit trajectories originating at S, the large difference in the final ψ values raises the question about which trajectory is conserving total energy. The following evaluation of the energy equation,

$$\frac{dT}{dt} = \frac{dK}{dt} + \frac{d\psi}{dt} - \frac{\partial\psi}{\partial t} = 0, \tag{13}$$

where

$$K = U^2/2,$$

shows that both techniques can yield different, yet energy-consistent, trajectories since $\partial\psi/\partial t$ is estimated differently for each formulation. In the explicit technique the finite form of (13) for the entire trajectory becomes

$$\left.\begin{aligned}
0 &= T_1 - T_0 = K_1 - K_0 + \psi_1 - \psi_0 - [\Delta\psi_1/(t_F - t_0)](t_1 - t_0)\\
0 &= T_2 - T_1 = K_2 - K_1 + \psi_2 - \psi_1 - [\Delta\psi_2/(t_F - t_0)](t_2 - t_1)\\
&\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\qquad \vdots\\
0 &= T_F - T_{F-1} = K_F - K_{F-1} + \psi_F - \psi_{F-1} - [\Delta\psi_F/(t_F - t_0)](t_F - t_{F-1})
\end{aligned}\right\}, \tag{14}$$

where $\Delta\psi_n$ is a local 12 h ψ change determined at the midpoint of the individual time increments $(t_n - t_{n-1})$. Summing (14) yields

$$0 = T_F - T_0 = K_F - K_0 + \psi_F - \psi_0 - \sum_{t=1}^{F} [\Delta\psi_t/(t_F - t_0)]\Delta t, \quad (15)$$

where $\Delta t = t_1 - t_0 = t_2 - t_1 = \ldots = t_F - t_{F-1}$. For the implicit scheme, the evaluation of (13) reduces to

$$0 = T_F - T_0 = K_F - K_0 + \psi_F - \psi_0 - (0.25 \times \Delta\psi_0 + 0.50 \times \Delta\psi_m + 0.25 \times \Delta\psi_F), \quad (16)$$

where $\partial\psi/\partial t$ is evaluated using the 12 h ψ tendencies at the beginning $(\Delta\psi_0)$, middle $(\Delta\psi_m)$ and end $(\Delta\psi_F)$ points of the trajectory. Even though T remains constant for (15) and (16), there is no constraint that the estimation of $\partial\psi/\partial t$ for the entire trajectory will be equal for both schemes. Therefore, K_F and ψ_F can be different for implicit and explicit trajectories beginning at the same point while still maintaining constant total energy.

For parcel S in Fig. 4, the explicit trajectory entered eastern Missouri where the ψ tendencies were noticeably positive. The parcel decelerated and continued turning to the right toward higher ψ values, remaining in regions of significantly positive $\partial\psi/\partial t$ throughout its entire journey. The explicit formulation accounted for the locally positive ψ tendencies that continuously contributed toward parcel deceleration along the entire trajectory. By comparison, the iteration between (2) and (16) for the implicit trajectory used one large positive ψ tendency at the beginning of the trajectory, but only slightly positive tendencies at the middle and end points. Therefore, the estimation of $\partial\psi/\partial t$ in (16) was smaller than the continuously accumulated local ψ changes in (15) for parcel S. As a result, the iteration between (2) and (16) yielded a trajectory ending near the North Carolina coast where ψ_F was significantly smaller than that obtained for the explicit trajectory ending in Alabama. This example illustrates an ability of the explicit scheme to objectively account for the effect that local ψ tendencies and geostrophic departures along the entire parcel path have on the calculated trajectory. This sensitivity of trajectories, which originate within regions of strong winds and large wind shears or pass through regions of significant local ψ tendencies, also illustrates the importance of careful ψ and wind analyses for computing trajectories with either the explicit or the implicit technique.

6. Summary

An explicit technique for calculating atmospheric trajectories is presented in this paper as an alternative method to the standard implicit scheme derived by Danielsen (1961). The technique uses the inviscid equations of motion and the discrete model formulation derived by Greenspan (1972, 1973) to compute trajectories on isentropic surfaces, assuming adiabatic flow. The discrete model formulation is designed specifically for a Lagrangian system and objectively accounts for the geostrophic departures, local ψ tendencies and subsequent accelerations along the entire length of the trajectory.

The initial test of the explicit technique used a simplified steady-state ψ distribution and various initial wind speeds to determine the magnitude of the errors introduced by the discretization of the equations of motion and by the grid point to parcel interpolations of the ψ gradient necessary to determine new accelerations at each time step. The discrete model approach yielded trajectories for which the change of total energy was two to three orders of magnitude less than the individual changes in either the kinetic or potential energies. The computed trajectories accurately simulated parcels undergoing an inertial oscillation in response to the initial ageostrophic component since parcels tested were within 50 m of their initial latitudinal position and 0.04 m s^{-1} of their initial wind speed after one full period of oscillation. The initial experiment demonstrated the stability of the discrete formulation and the accuracy of the interpolation scheme in maintaining zonal geostrophic trajectories and in accurately accounting for the accelerations resulting from the initially unbalanced flow.

The application of the discrete formulation to a diagnostic case study yielded favorable results. The trajectories ended with computed velocities very close to the observed winds at the 12 h mark. Comparisons with trajectories determined using Danielsen's implicit technique were also generally good. In general, trajectories computed with either technique ended within the error bounds expected from observation and analysis errors. However, significant differences occurred for trajectories initialized within the exit region of jet streaks where the influences of the geostrophic departures, local ψ tendencies and subsequent decelerations are very important. For these trajectories, the implicit scheme was evidently not always capable of objectively accounting for the effects of ageostrophic winds and local ψ tendencies along the entire trajectory and thus yielded significantly different results.

The application of the discrete model approach and the implicit scheme also revealed the sensitivity of the trajectories to the initial and final ψ fields and the initial wind analysis. Great care must be taken in the diagnostic analyses to achieve representative trajectories with either scheme. The use of the trajectories from case studies should also be

tempered by recognizing that the inviscid, adiabatic and linear ψ tendency assumptions are made for both the explicit and implicit schemes. While results from this case study are encouraging, the computation of trajectories for other cases should be attempted to further evaluate the application of this "discrete model" approach for calculating trajectories using a real data base.

Acknowledgments. We wish to thank Dr. Donald F. Greenspan for his inspiration and encouragement during this study, and Dr. Christopher Hayden and Mr. Tom Koehler for their critical review of the manuscript. We gladly thank the reviewers for their suggestions which helped clarify portions of the manuscript. We also express our appreciation to Mr. Gregory Krause for providing an independent, implicitly determined set of trajectories for the case study, and Mr. John Stremikis for drafting the figures.

This research was sponsored in part by the National Oceanic and Atmospheric Administration (Meteorological Satellite Laboratory) under Grant 01-4-158-2 and the National Science Foundation, under Grant ATM75-23223.

REFERENCES

Bleck, R., and P. L. Haagenson, 1968: Objective analysis on isentropic surfaces. NCAR Tech. Note NCAR-TN-39, Boulder, CO., 27 pp.

Buzzi, A., and R. Rizzi, 1975: Isentropic analyses of cyclogenesis in the lee of the Alps. *Riv. Ital. Geofis., Atti del XIII Cong. Int. Meteor., Alpina*, 1, 7–14.

Danielsen, E. F., 1961: Trajectories: isobaric, isentropic and actual. *J. Meteor.*, 18, 470–486.

——, 1966: Research in four-dimensional diagnosis of cyclonic storm cloud systems. Rep. No. 66-30, Air Force Cambridge Res. Lab., Dept. 66-30, Bedford, MA. 53 pp. [NTIS AD 632 668].

——, 1967: Moist isentropic flow and trajectories in a developing wave cyclone. Rep. No. 67-0617, Air Force Cambridge Res. Lab., Dept. 67-0617, Bedford, MA. 1–34. [NTIS AD 670 847].

——, 1968: Stratospheric–tropospheric exchange based on radioactivity, ozone and potential vorticity. *J. Atmos. Sci.*, 25, 502–518.

Greenspan, D., 1972: A new explicit discrete mechanics with applications. *J. Franklin Inst.*, 294, 231–240.

——, 1973: *Discrete Models*. Addison-Wesley, 165 pp. (see pp. 103–106).

Lilly, D. K., 1965: On the computational stability of numerical solutions of time-dependent non-linear geophysical fluid dynamics problems. *Mon Wea. Rev.*, 93, 11–26.

Newton, C. W., 1959: Axial velocity streaks in the jet stream: ageostrophic "inertial" oscillations. *J Meteor.*, 16, 638–645

Reap, R. M., 1972: An operational three-dimensional trajectory model. *J. Appl. Meteor.*, 11, 1193–1202.

Reiter, E. R., 1972: *Atmospheric Transport Processes. Part 3: Hydrodynamic Tracers*. AEC Critical Review Series, 206 pp.

Sechrist, F. S., and J. A. Dutton, 1970: Energy conversions in a developing cyclone. *Mon. Wea. Rev.*, 98, 354–362.

Shapiro, M. A., 1970: On the applicability of the geostrophic approximation to upper-level frontal-scale motions. *J. Atmos. Sci.*, 27, 408–420.

——, 1978: Further evidence of the mesoscale and turbulent structure of upper level jet stream-frontal zone systems. *Mon. Wea. Rev.*, 106, 1100–1111.

Whittaker, T. M., and R. A. Petersen, 1977: Objective cross-sectional analyses incorporating thermal enhancement of the observed winds. *Mon. Wea. Rev.*, 105, 147–153.

Young, J. A., 1968: Comparative properties of some time differencing schemes for linear and nonlinear oscillations. *Mon. Wea. Rev.* 96, 357–364.

N81-21692

Paper 51

NASA Technical Memorandum 80596 December 1979

Tropical Cyclone Rainfall Characteristics As Determined From a Satellite Passive Microwave Radiometer

Edward B. Rodgers and Robert F. Adler
Laboratory for Atmospheric Sciences (GLAS)
Goddard Space Flight Center
National Aeronautics and Space Administration
Greenbelt, MD 20771

ABSTRACT

Data from the Nimbus-5 Electrically Scanning Microwave Radiometer (ESMR-5) have been used to calculate latent heat release (LHR) and other rainfall parameters for over 70 satellite observations of 21 tropical cyclones during 1973, 1974, and 1975 in the tropical North Pacific Ocean. The results indicate that the ESMR-5 measurements can be useful in determining the rainfall characteristics of these storms and appear to be potentially useful in monitoring as well as predicting their intensity. The ESMR-5 derived total tropical cyclone rainfall estimates agree favorably with previous estimates for both the disturbance and typhoon stages. The mean typhoon rainfall rate ($1.9\ mm\ h^{-1}$) is approximately twice that of disturbances ($1.1\ mm\ h^{-1}$).

Case studies suggest that tropical cyclone intensification is indicated by the increase in the ESMR-5 derived LHR, the increase in the relative contribution of the heavier rain rates (≥ 5 mm h^{-1}) to the total storm rainfall, and the decrease in the radius of maximum rain rate from the cyclone center. It also appears evident from these case studies that by monitoring the trend of increasing LHR the first indication of tropical cyclone intensification may be obtained 1-2 days prior to the tropical cyclone reaching storm stage and often prior to the first reconnaissance aircraft observation. Further, the time of the maximum intensity of the tropical cyclone lags by 1-2 days the time of maximum LHR. The statistics of the Western Pacific tropical cyclones confirm the case study results in that tropical cyclone intensity can be monitored from ESMR-5 derived rainfall parameters. As the mean tropical cyclone intensifies from a disturbance to typhoon stage the average LHR increases steadily. The mean relative contribution of the heavier rain rate ($\geq 5\ mnh^{-1}$) to the total storm rainfall increased from .24 at depression stage to .33 at storm stage and finally to .39 at typhoon stage. The radial distance of the maximum rain rate from the center decreases with intensification while the azimuthal distribution indicates a slight preference for maximum rain rate in the right half of the composite storm at all stages. The study also indicates that Eastern Pacific hurricanes have less LHR, are more compact, and have less intense rainfall than the Western Pacific typhoons.

Paper 52

Reprinted from *Journal of Applied Meteorology*, 18, p. 978-991, 1979.

A Statistical Technique for Determining Rainfall over Land Employing Nimbus 6 ESMR Measurements

EDWARD RODGERS

NASA/GSFC, Greenbelt, MD 20771

HONNAPPA SIDDALINGAIAH[1]

Computer Sciences Corporation, Silver Spring, MD 20910

A.T.C. CHANG AND THOMAS WILHEIT

NASA/GSFC, Greenbelt, MD 20771

(Manuscript received 6 December 1978, in final form 10 May 1979)

ABSTRACT

At 37 GHz, the frequency at which the Nimbus 6 Electrically Scanning Microwave Radiometer (ESMR 6) measures upwelling radiance, it has been shown theoretically that the atmospheric scattering and the relative independence on electromagnetic polarization of the radiances emerging from hydrometeors make it possible to monitor remotely active rainfall over land. In order to verify experimentally these theoretical findings and to develop an algorithm to monitor rainfall over land, the digitized ESMR 6 measurements were examined statistically

Horizontally and vertically polarized brightness temperature pairs (T_H,T_V) from ESMR 6 were sampled for areas of rainfall over land as determined from the rain recording stations and the WSR 57 radar, and areas of wet and dry ground (whose thermodynamic temperatures were greater than 5°C) over the southeastern United States These three categories of brightness temperatures were found to be significantly different in the sense that the chances that the mean vectors of any two populations coincided were less than 1 in 100 Since these categories were significantly different, classification algorithms were then developed. Three decision rules were examined the Fisher linear classifier, the Bayesian quadratic classifier, and a non-parametric linear classifier The Bayesian algorithm was found to perform best, particularly at a higher confidence level An independent test case analysis showed that a rainfall area delineated by the Bayesian classifier coincided well with the synoptic-scale rainfall area mapped by ground recording rain data and radar echoes

1. Introduction

Precipitation is a fundamental meteorological parameter and it functions as an indicator, determinant or component of the distribution and amount of latent heat release which is critical to the understanding of storm and global atmospheric energetics and of the total hydrological cycle. The ability to monitor the coverage and movement of rain over land areas is important because of the direct impact of rain on crop production and also its influence on insect breeding areas and migration (Idso *et al.*, 1975). Moreover, the destructive effects due to heavy rainfall could be reduced by advance warnings furnished by satellites that map regions of heavy rain.

Since the advent of the polar orbiting and geosynchronous satellites, quantitative techniques have been developed to estimate rainfall indirectly. Estimations of rainfall have been made by correlating rain rate and amount with either cloud cover and type, cloud brightness or cloud temperature, utilizing visible and infrared sensors on board these satellites (Barnett, 1970, 1973; Martin and Scherer, 1973, Martin *et al.*, 1975; Follansbee and Oliver, 1975, Scofield and Oliver, 1977; Griffith *et al.*, 1978). However, all these techniques suffer from being only indirectly related to rainfall.

The microwave technique developed by Wilheit *et al* (1977) has a direct physical relationship with rain rate but only over ocean areas. This technique establishes a relationship between rain rate in the dynamic range of 1-20 mm h^{-1} and brightness temperatures (T_B) measured by the Electrically Scanning Microwave Radiometer on board Nimbus 5 (ESMR 5), which senses at 19.35 GHz upwelling radiation emitted by the earth and its atmosphere.

Meneely (1975) demonstrated that rainfall rate and coverage cannot be delineated using ESMR 5 mea-

[1] Present affiliation OAO Corp., Beltsville, MD 20705

0021-8952/79/080978-14$07.50

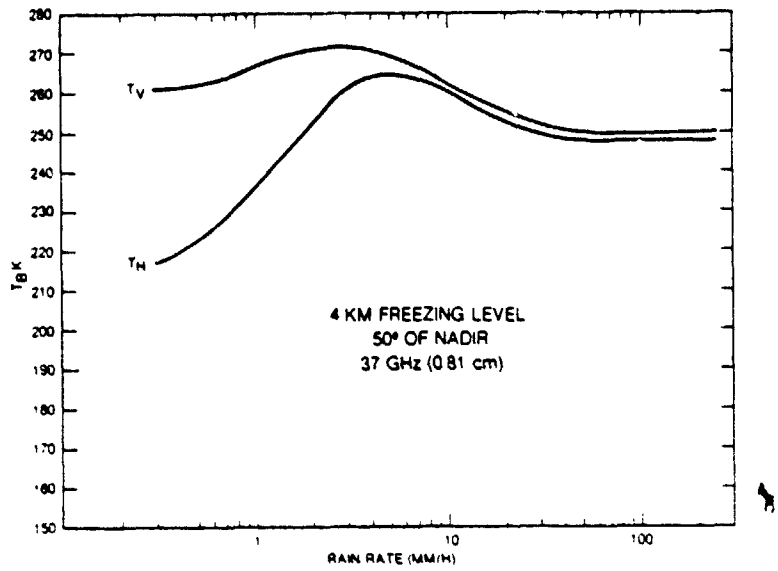


FIG. 1. Computed horizontally and vertically polarized brightness temperature at 37.0 GHz as a function of rain rate.

surements over land areas. This is because the rain has only a weak effect on the upwelling T_B from the land and the effect of soil moisture is comparable. Thus, although rain-like pattern can be discerned in the data, they correspond to both active rain areas and areas with moist soil. McFarland and Blanchard (1977), however, did demonstrate that rain amounts over land could be estimated indirectly by monitoring temporal changes in the ESMR 5 T_B.

Savage and Weinman (1975) and Savage *et al.* (1976) demonstrated theoretically that at 37.0 GHz (the frequency at which the Nimbus 6 ESMR sensor measures upwelling radiance) the scattering by hydrometeors is strong enough to provide a qualitative estimate of rain coverage over land. Furthermore, Weinman and Guetter (1977) demonstrated from a theoretical consideration that the upwelling radiation at 37.0 GHz emerging from rain clouds was essentially unpolarized and therefore was in contrast with the radiation emanating from wet surface background. According to the electromagnetic theory, if the emissivity of a surface is reduced by increasing its dielectric constant (as by adding moisture), then the emissivity will be highly polarized when the surface is viewed obliquely. These results are demonstrated in Fig. 1 which displays theoretically calculated bipolarized 37.0 GHz T_B at 40° incidence angle with the earth surface for a given rain rate. These T_B's were derived from a radiative transfer model with Lambertian reflection (Born and Wolf, 1975) from land surfaces at a thermodynamic temperature of 229.1 K and with a fixed dielectric constant and an atmospheric freezing level at 4 km (Wilheit *et al.*, 1977). It is seen from this figure that as rain rate increases beyond 4 mm h^{-1} T_B decreases due to strong backscattering by the large raindrops. Also the polarization difference becomes smaller. Moreover, Hall *et al.* (1978) inferred theoretically that information analogous to that provided by the National Weather Service radar summary charts can be produced when both ESMR 6 and the Temperature Humidity Infrared Radiometer (THIR) 11.5 μm data on board Nimbus 6 are used.

Thus, the sum and substance of these theoretical investigations is that the obliquely viewed 37 GHz radiation emitted by wet soil surfaces is polarized ($T_V > T_H$), whereas radiation emanating from dry land or heavy rainfall areas is essentially unpolarized ($T_V \approx T_H$). Moreover, T_B's upwelling from dry land areas are distinguishably higher than those from heavy rainfall areas or wet land surfaces. Hence, according to these theoretical conclusions rainfall over land can be at least qualitatively delineated and therefore its coverage and movement can be monitored irrespective of the land background by employing 37 GHz measurements from the ESMR on board Nimbus 6. Quantitative measurement of rainfall over land using a 37 GHz radiometer, however, appears less promising.

It is the purpose of this paper to substantiate the above conclusions and to arrive at an algorithm for the detection of rain over land by statistically analyzing ESMR 6 data. This statistical analysis will be performed by first sampling three categories of ESMR 6 T_B's (representing rain over land, wet land surfaces without rain and dry land surfaces), then testing these populations for uniqueness and separability, and finally developing a classification algorithm to delineate rain over land.

TABLE 1. Dates of synoptic rain cases used to develop ESMR 6 classification algorithms.

Case	Date	Time (GMT)
1	31 Jul 1975	1620
2	4 Aug 1975	1635
3	1 Oct 1975	1700
4	7 Nov 1975	1700
5	12 Nov 1975	1700
6	29 Dec 1975	1717
7	3 Jan 1976	1715
8	6 Jan 1976	1635

2. The ESMR 6 system

The ESMR 6 system flown aboard Nimbus 6 (Wilheit, 1975) receives the thermal radiation upwelling from the earth's surface and atmosphere in a 250 MHz band centered at 37 GHz. The antenna beam scans electrically an arc of 70° in 71 steps ahead of the spacecraft along a conical surface with a constant earth incidence angle of 40° every 5.3 s. The nominal resolution is 20 km crosstrack and 45 km downtrack. The instrument measures both horizontal and vertical polarization components by using two separate radiometric channels. The data are calibrated using warm (instrument ambient) and cold (cosmic background) inputs to the radiometer.

Examination of the data revealed two problems affecting calibration, apparent modulation of the loss of the antenna during the orbit and excess noise when measuring the radiation from the warm calibration load. The modulation of the antenna loss was found to be consistent with respect to sun angle. It showed its maximum rate of change as the spacecraft entered the sunlight with the instrument facing the sun and fairly rapid changes whenever the spacecraft was in the sunlight. The changes were rather gradual whenever the spacecraft was in darkness. The most reasonable explanation seems to be thermomechanical warping of the radome modulating the coupling among the radiating elements in the antenna. An empirically derived correction as a function of beam position has been applied to the data to correct the biases in the data due to antenna losses. No cause is easily discerned for excess noise in the warm calibration load However, the data were rejected whenever the warm calibration load was too high (>310 K) or too low (<290 K). Otherwise, an empirical correction was applied to the data to mitigate excessive noise due to warm calibration load. It was found in a more subtle study by Wilheit (1978), where the effects of wind on ocean surface emission at 37 GHz were estimated, that the residual errors of these problems were excessive whenever the instrument was in the sun. Therefore, only nighttime data were used for that study.

The T_B as observed from the satellite is dependent upon the emission from the earth's surface modified by the intervening atmosphere. The emissivity, a function of the dielectric constant, is variable over land surfaces (depending on vegetation, soil type, soil moisture, etc.) and generally is large (~0.9). In rain situations three constituents contribute significantly to the absorption: molecular oxygen (Meeks and Lilley, 1963), water vapor (Staelin, 1966) and liquid water droplets (Mie, 1908; Gunn and East, 1954). Water droplets contribute more significantly to absorption and reemittance than the other constituents and are the only source of scattering at this frequency. Ice crystals are essentially transparent at this frequency.

3. Data sampling

Simultaneous ground station and radar measurements of rain and ESMR 6 T_B were needed in order to develop an algorithm which classified a given ESMR 6 instantaneous field of view (IFOV) as rain over land, dry land surface or wet land surface. Eight daytime synoptic-scale rainfall cases over the southeastern United States were used where surface rainrate data taken from stations reporting hourly rainfall amounts and from the WSR 57 radar coincided with Nimbus 6 overpass to within 5 min. The surface temperature in each of these cases was not less than 5°C. Rain areas were sampled within areas delineated as rain by either the WSR 57 radar (rain rates ≥ 2.5 mm h^{-1}) and/or the stations reporting hourly rainfall amounts. The dates and time of the occurrence of these cases are given in Table 1. Wet land surfaces were sampled upwind and adjacent to the raincells observed on the WSR 57 radar and dry land surfaces were sampled over areas where rain had not fallen within a 24 h period previous to the Nimbus 6 pass.

Fig. 2 illustrates the sampling technique. The figure shows the ESMR 6 horizontally polarized T_B's (K) measured at approximately 1655 GMT 6 January 1976 together with rainfall as delineated by the WSR 57 radar (located at Waycross, Georgia at 1700 GMT) and by stations reporting hourly rainfall

TABLE 2. Elementary statistics of sampled data (surface temperature ≥5°C).

Sample size (N)	Rain area 216		Dry ground 189		Wet soil 66	
	T_{BR}	T_{VR}	T_{BD}	T_{VD}	T_{BW}	T_{VW}
Mean (μ)	254.53	260.98	271.46	278.18	252.05	268.86
Mean brightness temperature difference	6.45		6.72		16.81	
Standard deviation (σ)	7.21	5.81	6.18	7.20	9.41	7.64
Sample correlation coefficient between T_B and T_V (ρ)	0.55		0.37		0.82	

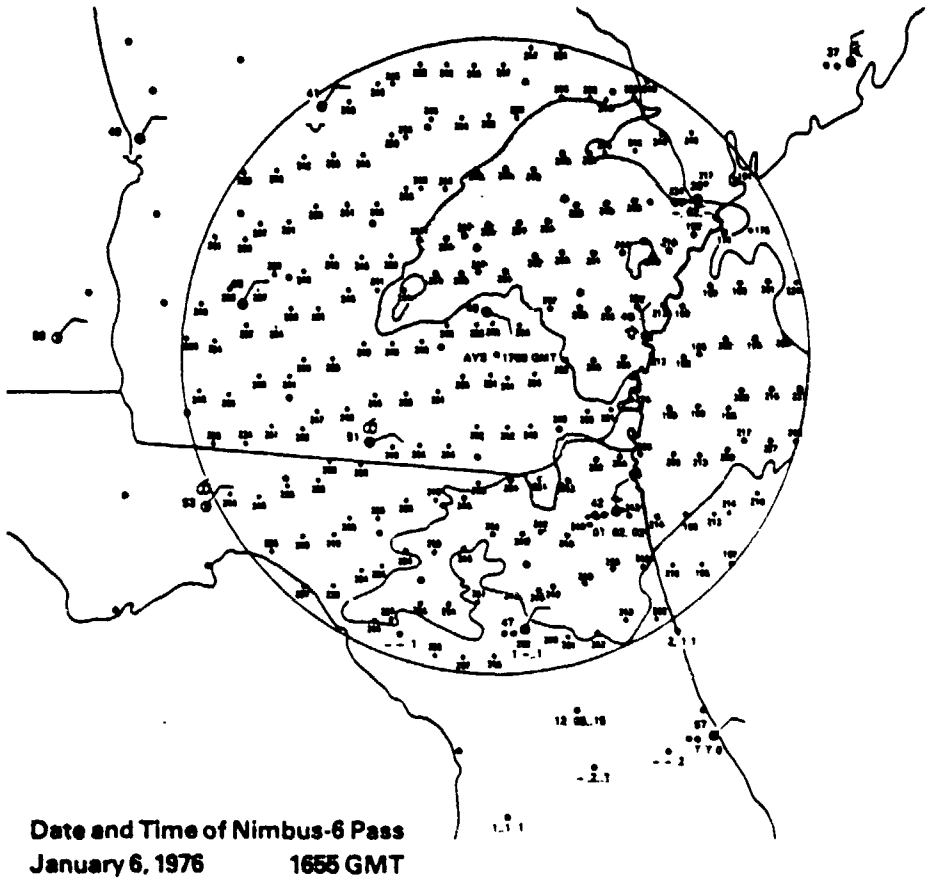


FIG. 2. ESMR 6 horizontally polarized T_B (1655 GMT 6 January 1976) superimposed on the PPI WSR 57 radar image at Waycross, Georgia (1700 GMT 6 January 1976)

amounts. The ESMR 6 T_B's are within the field of view of the radar where the circle shows the outer bounds of the PPI image at a 232 km radius. The shaded area represents rain (rain rate ≥ 2.5 mm h^{-1}). The large dots are hourly rain recording stations where rain amounts (in inches) for hours ending at 1700, 1600 and 1500 GMT are displayed according to model in the figure. If no rain has fallen during that period, no measurements are shown. Station models reporting temperature, present weather, cloud type and amount, and wind direction and speed for 1800 GMT are also given. The small dots locate the center of the ESMR 6 footprints. For this case, ESMR 6 T_B's representing rain over land was sampled within the shaded area. The T_B's representing wet land surfaces were sampled southwest of the shaded area since the rain area was moving northeast, and T_B's representing dry land surfaces were sampled over western Georgia where rain had not fallen within 24 h of the Nimbus 6 pass. It should be noted that the horizontally polarized T_B's over the radar echoes and the wet or dry land areas outside the radar echoes are all about the same. The reason for this non-variability of T_B's is that the rain in this case was light and did not significantly influence the horizontally polarized T_B's at 37 GHz.

4. Statistical analysis

Elementary statistics of the total sampled data (ESMR 6 measurements where surface thermodynamic temperatures were greater than 5°C) are presented in Table 2. The table gives for each category the sample size, the mean and standard deviation of the horizontally and vertically polarized T_B, and the correlation and the mean difference between horizontally and vertically polarized T_B's. These data are also shown as a scatter plot in Fig. 3. In this figure the C represents the mean of the population and each frequency concentration ellipse encompasses 68% (one

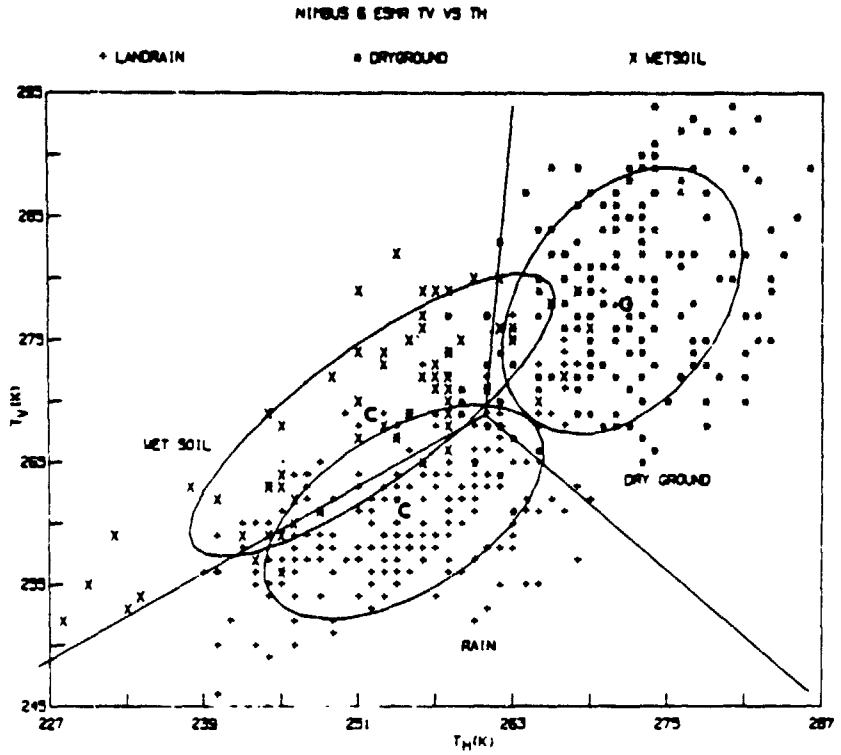


FIG. 3. Vertically polarized vs horizontally polarized EMSR 6 T_B for each sampled category (rain over land, and wet and dry land surfaces)

standard deviation) of the data within the population. The ellipses reveal the extent of scattering of data from each population, the correlation between the dual polarization T_B's (T_H and T_V) within each population (the higher the correlation the larger the eccentricity of the ellipse), and the extent of overlap among the populations. The three concurrent lines drawn in this figure are the Fisher (1938) linear discriminant lines which separate two-by-two the rain over land area (S_R), the dry land surface (S_D), and the wet land surface (S_W) populations represented by the T_B pairs (T_H,T_V).

It can be seen from Fig. 3 and Table 2, that T_B's from rain areas over land are colder than those T_B's from dry land surface areas. Further, the difference between the mean horizontally and vertically polarized T_B's from rain areas over land (6.45 K) is much smaller than that for wet land surfaces (16.81 K). This is in accordance with theoretical findings that microwave radiation emerging from hydrometeors is essentially unpolarized (Weinman and Guetter, 1977) whereas radiation emanating from wet land surfaces is polarized. It is also seen from Fig. 3 that the largest overlap occurs between the data obtained from rainfall areas and wet land surfaces. The reason for this is that sometimes in sampling rain over land the total upwelling radiance received by the radiometer contains a direct surface contribution. This may occur when an IFOV of the ESMR 6 measurement is partially filled with moderate to heavy rain or when it is completely filled with light rain (background being wet land surface). Consequently, the T_B's for each category are somewhat similar, thus producing the overlap between rain over land and wet land surface classes.

Since the surface emission is given by ϵT_S, where ϵ is the surface emissivity and T_S the surface thermodynamic temperature, there is an influence of T_S on ESMR 6 measured dry land surface T_B. A decrease in T_S results in a decrease in T_B from dry ground and consequently, the T_B contrast between dry land surfaces and rain over land will also decrease. These effects can be observed in Figs. 4 and 5 and in statistical Tables 3 and 4. The figures and tables are identical to Fig. 3 and Table 2, respectively, except that Fig. 4 and Table 3 correspond to sample cases where the surface thermodynamic temperatures were above 15°C, while Fig. 5 and Table 4 correspond to cases where the surface thermodynamic temperatures were between 5° and 15°C. It is clear from Fig. 5 and Table 4 that rain over land is difficult to delineate from dry land surfaces when the surface thermodynamic temperature is below 15°C. Since the populations in Fig. 5 cannot be separated, the Fisher linear discriminant lines are not drawn.

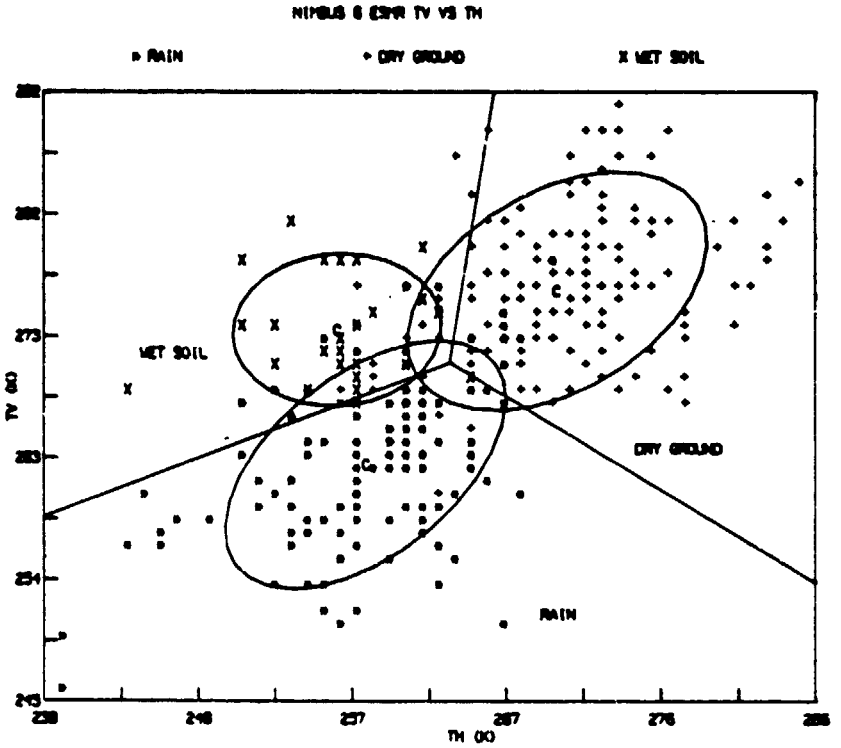


FIG. 4. As in Fig. 3 except for surfaces whose thermodynamic temperatures are greater than 15°C.

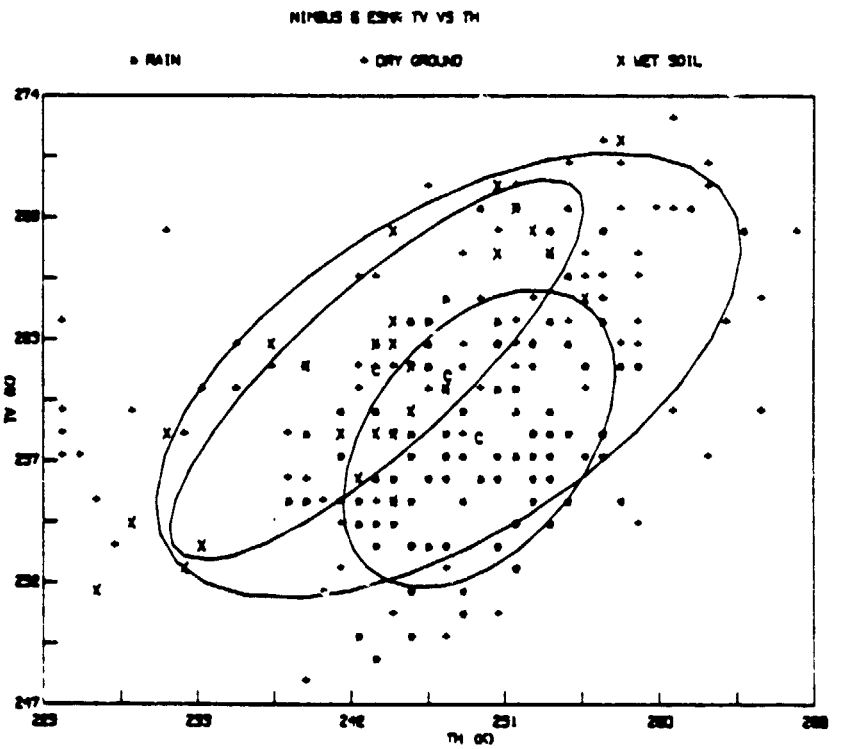


FIG. 5. As in Fig. 3 except for surfaces whose thermodynamic temperatures are between 5° and 15°C.

TABLE 3. Elementary statistics of sampled data (surface temperature >15°C).

Sample size (N)	Rain area 112		Dry ground 145		Wet soil 26	
	T_{BR}	T_{VR}	T_{BD}	T_{VD}	T_{BW}	T_{VW}
Mean (μ)	258.57	263.22	270.22	276.61	256.81	273.62
Mean brightness temperature difference	4.65		6.39		16.81	
Standard deviation (σ)	5.70	6.36	6.07	6.16	4.22	3.90
Sample correlation coefficient between T_B and T_V (ρ)	0.53		0.62		0.08	

Fig. 6 displays the marginal densities (histograms) of the sampled horizontally and vertically polarized T_B's from the three populations. Table 5 presents the results of the chi-square test (Cochran, 1952) performed to validate the normal distribution of the data. Since each observed chi-square value in Table 5 is comparable to the corresponding critical (table) value at 1%, it is assumed that each marginal distribution of the data is Gaussian. Therefore, it is reasonable to assume that the data from each of the populations S_R, S_D or S_W satisfy the bivariate Gaussian density distribution

$$f(\mathbf{x}) = \frac{1}{2\pi|c|^{\frac{1}{2}}} \exp[-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T c^{-1}(\mathbf{x}-\boldsymbol{\mu})], \quad (1)$$

where $\mathbf{x}$ is the two-dimensional column vector (T_H, T_V), $\boldsymbol{\mu}$ is the mean of $\mathbf{x}$, c is the covariance matrix of the population, c^{-1} is the inverse of c, $|c|$ is the determinant of c, and $(\mathbf{x}-\boldsymbol{\mu})^T$ is the transpose of $(\mathbf{x}-\boldsymbol{\mu})$. The $\boldsymbol{\mu}$ and c are estimated using the sampled data from each class. Then $\boldsymbol{\mu}$'s are provided by Table 2 and c, c^{-1} and $|c|$ by Table 6.

Prior to employing the data in Table 2 for the purpose of developing classification algorithms, the data were examined to verify whether the three populations were statistically distinguishable from one another. To accomplish this, an F (variance ratio) test, in terms of Hotelling's T^2 and Mahalanobis's D^2 (Kshirsagar, 1972), was performed to determine the

TABLE 4. Elementary statistics of sampled data (surface temperature 5–15°C).

Sample size (N)	Rain area 104		Dry ground 98		Wet soil 28	
	T_{BR}	T_{VR}	T_{BD}	T_{VD}	T_{BW}	T_{VW}
Mean (μ)	249.92	258.78	248.11	261.57	244.04	261.82
Mean brightness temperature difference	8.86		13.46		17.78	
Standard deviation (σ)	5.16	4.40	11.08	6.58	7.87	5.59
Sample correlation coefficient between T_B and T_V (ρ)	0.43		0.58		0.84	

TABLE 5. Chi-square test for normality.

	Rain		Dry		Wet	
	T_B	T_V	T_B	T_V	T_B	T_V
Number of cells	8	8	8	9	6	5
Degrees of freedom	5	5	5	6	3	2
Table value of χ^2 at 0.01	15.09	15.09	15.09	16.81	11.34	9.21
Observed value of χ^2	14.57	28.17	18.99	11.59	8.93	10.33

significance of the differences between the means of any two classes. Then the simultaneous confidence intervals were estimated for these differences by Scheffé's procedure (Scheffé, 1959; Bennett, 1951).

Table 7 displays D^2 and T^2 as well as the computed and table (critical) values of F. The difference between the means of any two classes is highly significant since the observed value of F is much higher for each pair of classes than the corresponding critical (table) value of F at the 1% confidence level, i.e., the probability that the mean vectors of any two populations are identical is less than 1 in 100.

Table 8 shows the estimated confidence intervals. It can be seen that only the interval for the differences between the wet land surface and rainfall over land mean horizontal polarization T_B's contains zero. Therefore, the three populations are distinguishable from one another when the dual polarization information is taken into consideration. However, the lower bounds of the mean differences between rainfall over land and wet land surface T_B's is smaller than those of the other two pairs. This indicates that it will be more difficult to distinguish an area of rain over land from wet land surfaces.

TABLE 6. Covariance matrices of sampled data.

Rain area

	Matrix		Inverse	
	T_{BR}	T_{VR}	T_{BR}	T_{VR}
T_{BR}	52.23	23.02	0.0273	−0.0185
T_{VR}	23.02	33.93	−0.0185	0.0420

Covariance matrix determinant: 1242.4240

Dry ground

	Matrix		Inverse	
	T_{BD}	T_{VD}	T_{BD}	T_{VD}
T_{BD}	38.36	16.51	0.030	−0.010
T_{VD}	16.51	52.14	−0.010	0.022

Covariance matrix determinant: 1727.7496

Wet soil

	Matrix		Inverse	
	T_{BW}	T_{VW}	T_{BW}	T_{VW}
T_{BW}	90.39	59.73	0.034	−0.035
T_{VW}	59.73	58.28	−0.035	0.053

Covariance matrix determinant: 1700.3525

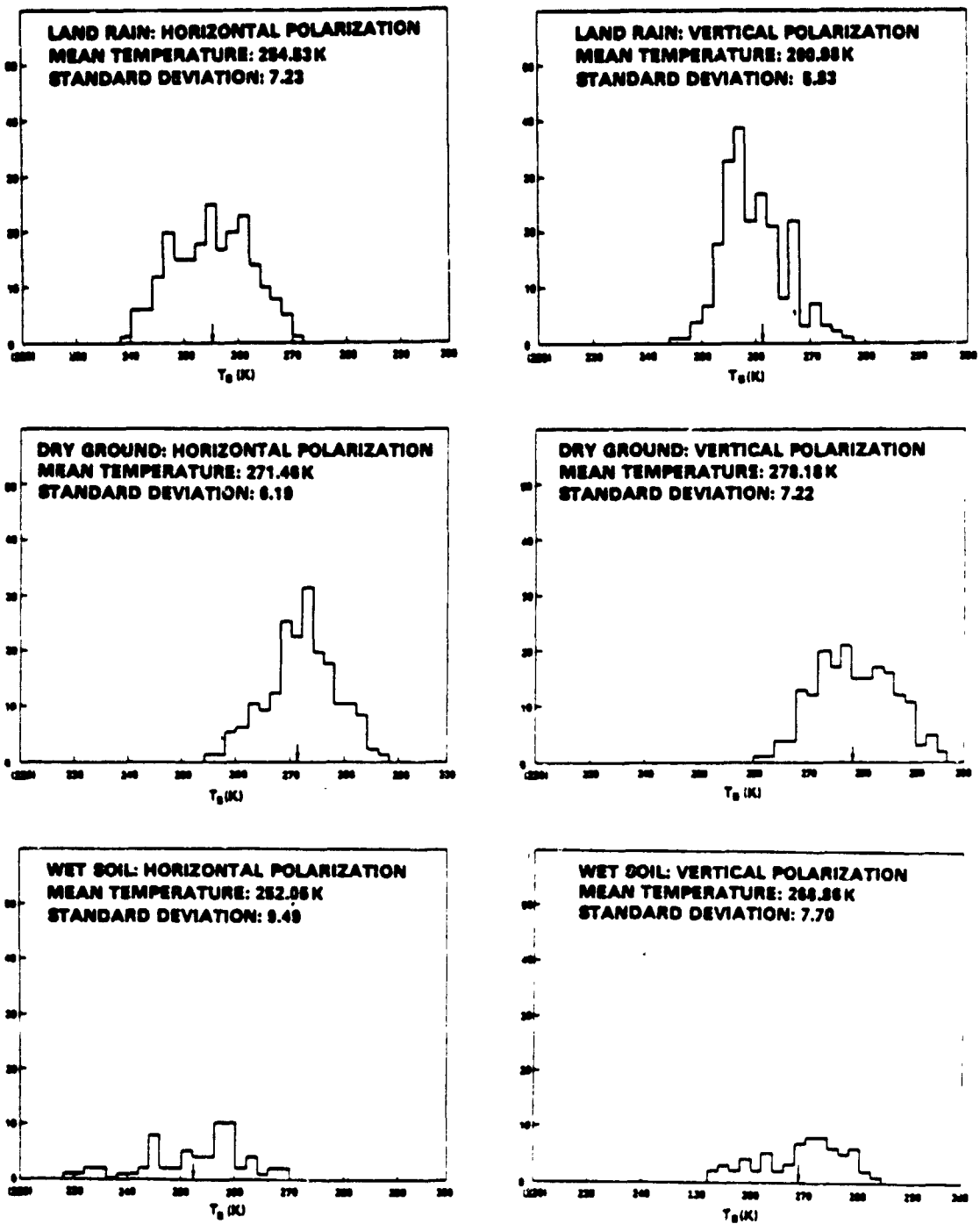


FIG. 6. Marginal densities (histograms) of the sampled horizontally and vertically polarized T_B's from the three populations.

5. Classification algorithms

Since the populations were found to be statistically distinguishable and satisfied the Gaussian frequency distribution, three classification techniques were considered with the purpose of developing an efficient and effective classification algorithm to detect and delineate active rainfall over land from dry and wet land surfaces. The three techniques are the Bayesian classifier, the Fisher linear discriminant classifier and a non-parametric linear discriminant classifier.

TABLE 7. Significance between means (F test).

	Mahalanobis's distance squared D^2	Hotelling's T^2	Observed variance ratio F	Table value of F at 1%
Rain vs dry	9.13	920.35	459.04	3.83
Dry vs wet	6.03	295.13	146.28	3.87
Rain vs wet	4.00	202.06	100.67	3.86

The Bayesian classifier is a parametric classifier (i.e., it assumes the functional form of the relevant density function). The non-parametric linear discriminant classifier does not assume a density function (Bond and Atkinson, 1972). The Fisher linear discriminant classifier may be either parametric or non-parametric (Fisher, 1938). All three methods are termed supervised in the sense that it is necessary to use known sample data for the various classes to train the algorithms. Algorithms were developed using all three classifiers and tested using independent data. It was found that the results from the Bayesian classifier were superior to the other two methods. Hence, only the Bayesian classification technique will be described in the following.

The Bayesian classifier is a Gaussian parametric maximum likelihood quadratic classifier which requires the knowledge of the *a priori* probabilities for the occurrence of each class (Duda and Hart, 1973; Fu *et al.*, 1969). It minimizes the average loss due to misclassification while assuming that each misclassification is equally costly.

It minimizes the conditional average loss

$$L(\mathbf{x},S_k)=\sum_{i=1}^{3}\lambda(S_k|S_i)P(S_i|\mathbf{x}), \tag{2}$$

where $\lambda(S_k|S_i)$ is the loss incurred when a measurement $\mathbf{x}=(T_H,T_V)$ actually belonging to class S_k is placed in class S_i and $P(S_i|\mathbf{x})$ is the *a priori* probability of the class S_i occurring having observed $\mathbf{x}$. The symmetrical loss function $\lambda(S_k|S_i)$ is given by

$$\lambda(S_k|S_i)=\begin{cases}0, & i=k\\ 1, & i\neq k\end{cases} \qquad i,k=1,2,3. \tag{3}$$

Hence, all misclassifications are equally costly and Eq. (2) now reduces to

$$L(\mathbf{x},S_k)=1-P(S_k|\mathbf{x}), \tag{4}$$

TABLE 8. Simultaneous confidence intervals for differences between mean brightness temperatures representing rain (R), dry (D) and wet (W) areas.

Polarization	Rain vs dry	Dry vs wet	Wet vs rain
Horizontal	$14.88 \leqslant \mu_{HD}-\mu_{HR} \leqslant 18.98$	$15.89 \leqslant \mu_{HD}-\mu_{HW} \leqslant 22.93$	$-1.15 \leqslant \mu_{HR}-\mu_{HW} \leqslant 6.11$
Vertical	$15.00 \leqslant \mu_{VD}-\mu_{VR} \leqslant 19.20$	$6.04 \leqslant \mu_{VD}-\mu_{VW} \leqslant 12.60$	$4.93 \leqslant \mu_{VW}-\mu_{VR} \leqslant 10.83$

where $P(S_k|\mathbf{x})$ is the conditional probability that the class S_k to which $\mathbf{x}$ is assigned is correct.

The likelihood function $P(S_k|\mathbf{x})$ is given by Bayes' rule

$$P(S_k|\mathbf{x})=P(\mathbf{x}|S_k)P(S_k)/\sum_{i=1}^{3}P(\mathbf{x}|S_i)P(S_i), \tag{5}$$

where $P(\mathbf{x}|S_i)$ is the bivariate Gaussian probability density function of $\mathbf{x}$ given that $\mathbf{x}$ is in S_i, and $P(S_i)$ is the *a priori* probability of the class S_i occurring. Sample data sizes given in Table 2 provide the values of $P(S_i)$. They are 0.459, 0.401 and 0.140 for the classes S_R, S_D and S_W, respectively.

Since the loss given by Eq. (4) is to be minimized, the quadratic discriminant functions are

$$g_i(\mathbf{x})=P(S_i)P(\mathbf{x}|S_i), \quad i=1,2,3. \tag{6}$$

These functions, considering the relationship in Eq. (1), lead to the following decision rule. The measurement $\mathbf{x}$ belongs to the class S_k if

$$2\ln P(S_k)-\ln|c_k|-(\mathbf{x}-\boldsymbol{\mu}_k)^T c_k^{-1}(\mathbf{x}-\boldsymbol{\mu}_k) > 2\ln P(S_i)-\ln|c_i|-(\mathbf{x}-\boldsymbol{\mu}_i)^T c_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) \tag{7}$$

for all $i\neq k$, where c_i and $\boldsymbol{\mu}_i$ are the covariance matrix and the mean vector of the class S_i. Then by substituting the relevant values into Eq. (7), one arrives at the following Bayesian algorithm. The pixel corresponding to the given vector (T_H,T_V) is rainfall over land, dry land surface or wet land surface, respectively, depending on which of the following values is the largest:

$$P_R(T_H,T_V)=-0.027T_H^2+0.038T_HT_V-0.042T_V^2+3.826T_H+12.250T_V-2094.097, \tag{8}$$

$$P_D(T_H,T_V)=-0.030T_H^2+0.020T_HT_V-0.022T_V^2+10.720T_H+6.811T_V-2412.165, \tag{9}$$

$$P_W(T_H,T_V)=-0.034T_H^2+0.070T_HT_V-0.053T_V^2-1.678T_H+10.846T_V-1261.721. \tag{10}$$

The quadratic function (Mahalanobis's distance squared)

$$Q_k(\mathbf{x})=(\mathbf{x}-\boldsymbol{\mu}_k)^T c_k^{-1}(\mathbf{x}-\boldsymbol{\mu}_k) \tag{11}$$

is a measure of distance in probabilistic terms between $\boldsymbol{\mu}_k$ and $\mathbf{x}$ and has a chi-square distribution with two degrees of freedom (Scheffe, 1959). Therefore a con-

TABLE 9. Probabilities of misclassification: Theoretical computation.

Known	Classified Rain	Dry	Wet
Rain	77.15	6.66	16.19
Dry	6.67	82.08	11.25
Wet	16.28	11.29	72.43

Average accuracy: 77.22%.

TABLE 10. Bayesian classification error matrix determined from sampled data.

Known	Classified Rain	Dry	Wet
Rain	89.35	6.02	4.63
Dry	7.41	91.53	1.06
Wet	27.27	15.15	57.58

Average accuracy: 79.49%.

fidence value F can be associated with each classified pixel. F is given by

$$F(\mathbf{x}) = 255\left\{1 - \frac{[Q_R(\mathbf{x})]^2}{n_\sigma}\right\}, \qquad (12)$$

where n_σ is the distance, in terms of standard deviation from the mean, to which zero confidence value is assigned.

This maximum likelihood decision rule selects one class from a set of predetermined classes (S_R, S_D, S_W in the present context) to which a pixel represented by $\mathbf{x} \equiv (T_H, T_V)$ most likely belongs. The associated confidence value, $F(\mathbf{x})$, measuring the distance in probabilistic terms of $\mathbf{x}$ from the mean point of the selected class, is calculated according to Eq. (12). The hypothesis that a pixel actually belongs to a selected class may be accepted or rejected based on this confidence value. If the confidence value is greater than a predetermined value (153.0, 178.5 and 204.0, respectively, in the case of 60, 70 and 80% confidence levels), the hypothesis is accepted and the pixel is put in the selected class; otherwise, the hypothesis is rejected and the pixel is put in the unknown class. The problem, of course, is to assign the appropriate values for thresholding the confidence value parameter.

6. Error analysis

An error estimate was made in order to evaluate quantitatively the performance of the Bayesian classification algorithm. The error rates were computed according to the asymptotic formulas given by Okamoto (1963), assuming that the populations satisfy the Gaussian distribution, have different means, and have the same covariance matrices. The results are shown in Table 9. Virtually all of the misclassification probability in each case was accounted for by the first term of the asymptotic expansion

$$\Phi(-\tfrac{1}{2}\Delta) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{-\frac{1}{2}\Delta} \exp\left[\frac{-z^2}{2}\right]dz, \qquad (13)$$

where

$$\Delta^2 = \frac{N_1+N_2-5}{N_1+N_2-2}D^2 - \frac{2(N_1+N_2)}{N_1N_2} \qquad (14)$$

D is Mahalanobis's distance and N_1, N_2 are sample sizes of the populations under consideration. Only a small fraction is contributed by the rest of the terms. From Table 9 it is clear that the chance of incorrectly classifying wet land surfaces or dry land surfaces as rain over land is nearly 23%. But when a given pixel is classified as a raining area and each of the eight contiguous pixels that cluster around it is also classified as rain over land, then the chance of misclassification of that central pixel is reduced to 7.7×10^{-6}% assuming each pixel is independently classified.

Table 10 displays the actual probabilities that the Bayesian algorithm classify the sampled training data into the various populations is as indicated. The average accuracy is the mean of the diagonal elements of the corresponding error matrix, and these averages compare well with the estimated average.

Tables 11 and 12 show the estimated error matrices corresponding to data which came from land areas where the surface thermodynamic temperature was less than or greater than 15°C, respectively. It is apparent from the tables that the classifications are not definitive when the surface thermodynamic temperature is between 5° and 15°C.

7. Algorithm evaluation

A case not previously used in sampling was tested to verify qualitatively the performance of the Bayesian classification algorithm. This case consisted of a synoptic-scale rain pattern over the southeastern United States (14 September 1976) which was observed by the ESMR 6 sensor (surface thermodynamic temperature $\geq$15°C). Fig. 7 shows the rainfall area delineated by the WSR 57 radars and hourly rainfall reporting stations. The approximate time of the radar PPI images was 1630 GMT (within 5 min of the Nimbus 6 pass). The reporting times of the hourly

TABLE 11. Probabilities of misclassification: Theoretical computation (<15°C).

Known	Classified Rain	Dry	Wet
Rain	45.99	36.23	17.78
Dry	36.26	23.99	39.75
Wet	18.03	41.86	40.11

Average accuracy: 36.70%.

TABLE 12. Probabilities of misclassification. Theoretical computation (>15°C).

Known	Classified Rain	Dry	Wet
Rain	74.83	11.76	13.41
Dry	11.75	77.66	10.59
Wet	13.59	10.72	75.69

Average accuracy: 76.06%.

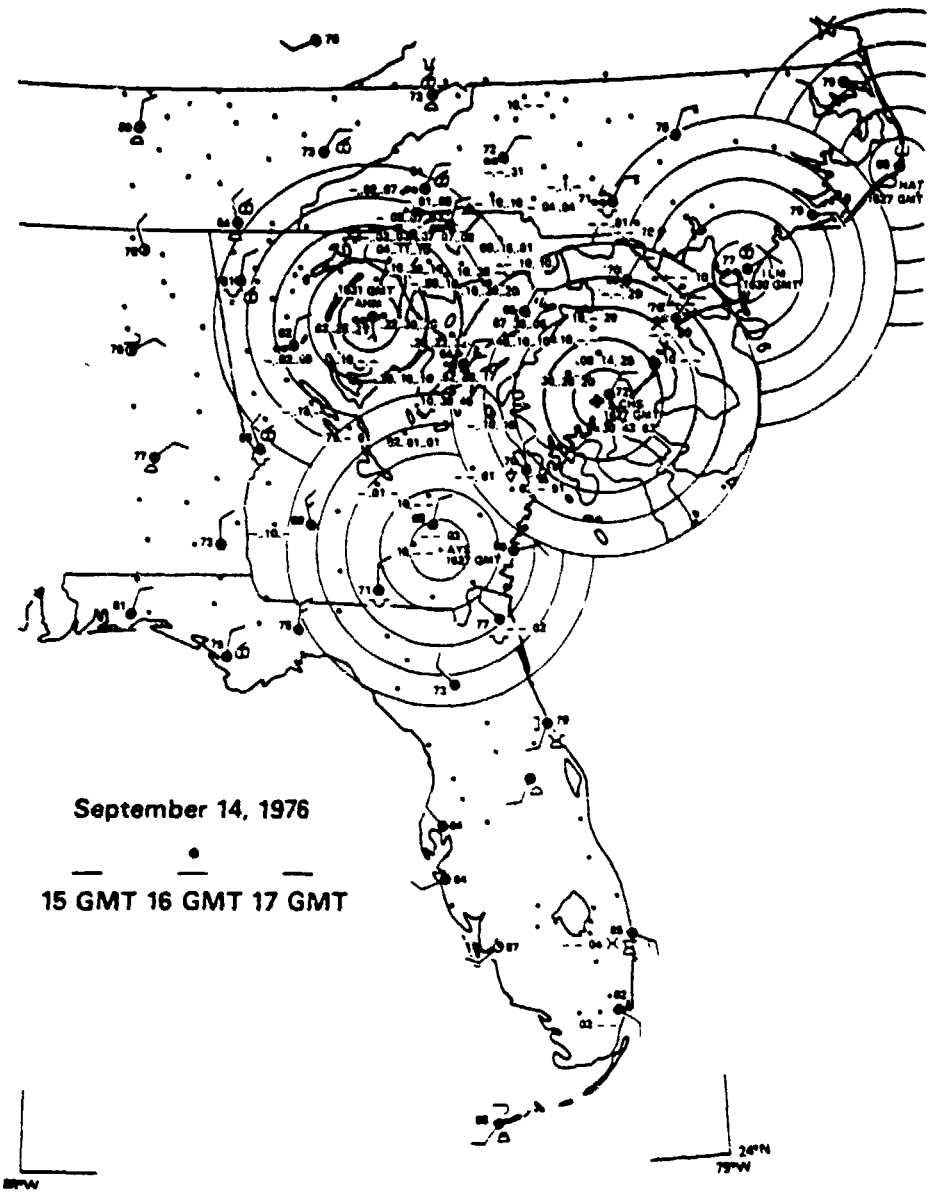


FIG. 7. Rainfall over the southeast United States as delineated by the WSR 57 radar and hourly rainfall reporting stations. Time of the data is approximately 1630 GMT 14 September 1976.

precipitation amounts were 1500, 1600 and 1700 GMT. The shaded area within the WSR 57 radar PPI range (232 km) is rainfall area with rain rates >2.5 mm h^{-1}. The radars were located at Waycross and Macon, Georgia; Charleston, South Carolina; and Wilmington and Cape Hatteras, North Carolina. Surface station data (present weather, temperatures, cloud type and amount, precipitation amount in three hours, and wind velocity and direction) were taken at 1800 GMT. Hourly rainfall is also shown. (See model in Fig. 7.)

The Bayesian (70 and 80% confidence) classification maps are seen in Figs. 8 and 9, respectively. Areas of clouds most likely producing rain are delineated by the Nimbus 6 THIR 11.5 μm channel where equivalent blackbody temperatures $T_{BB} \leq 270$ K (Shenk *et al.*, 1976). Rain areas in the absence of rain producing clouds are considered misclassifications. Regions only covered by clusters of contiguous pixels classified into a single individual class are shown, since the probability of misclassifying clusters is much less than that of a single pixel.

It is seen by comparing the two Bayesian classification maps at 70 and 80% confidence level (Figs. 8 and 9, respectively) with the map delineating observed rain (Fig. 7) that they agree well, particularly at the 80% confidence level. No attempt was made to verify

wet land surfaces. The 80% confidence Bayesian classification, however, did not delineate rain over eastern South Carolina as well as the 70% confidence classification. This is because the requirement of 80% confidence level is obviously too stringent.

The main discrepancies found between the ESMR 6 observed rainfall and ground observed rainfall is seen over North Carolina and southwestern Georgia. The rainfall indicated by ESMR 6 over North Carolina may be suspended liquid water in the clouds and/or virga ahead of the rain area (the area of rain was moving northeastward toward North Carolina). The ESMR 6 delineated rain over southwestern Georgia, which was upstream from the rain area, may be due to wet land surfaces produced by the rain that fell a few hours prior to the Nimbus 6 pass.

The Bayesian classification algorithm was applied to another test case (1645 GMT 27 August 1976, surface thermodynamic temperatures were $\geq 15°C$) over the same geographical area as the previous case in order to determine whether the surface characteristics (vegetation, soil moisture and surface roughness) had influenced the classification performed in the previous case. During these periods the area was under the influence of a Bermuda high. Also there was a squall line located in southwestern Virginia and extending southwestward into Tennessee. Rainfall

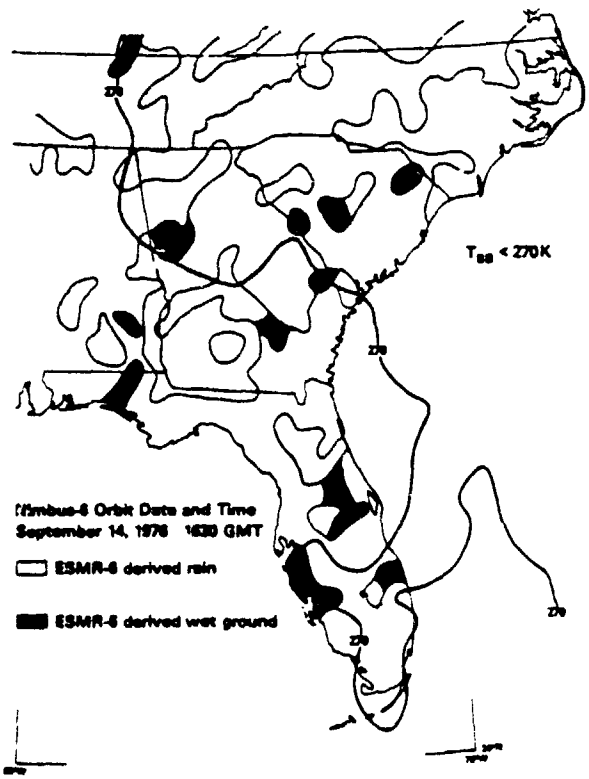


FIG. 8. ESMR 6 derived rainfall distribution using the Bayesian classifier with a confidence level of 70%. Time of Nimbus 6 pass was 1630 GMT 14 September 1976.

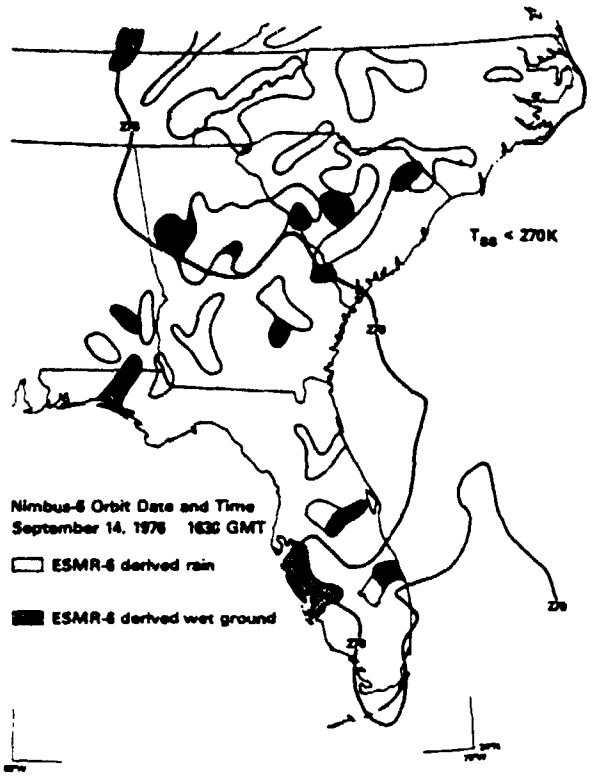


FIG. 9. As in Fig. 8 except with a confidence level of 80%.

was associated with this squal line and along the Gulf States. Fig. 10 shows the 80% confidence level Bayesian classification map superimposed over the surface station models. The reporting time for these stations was 1800 GMT. The figure shows that the only areas classified as rain over land were along the Gulf Coast and in eastern Tennessee. The regions in the previous case where the algorithm showed rainfall were classified as dry land surfaces. Hence, there were no influences of extraneous surface characteristics on the outcome of the previous case study.

However, contradicting results occurred when the Bayesian classification algorithm was applied to a nighttime Nimbus 6 pass over the same geographical area (0525 GMT, 13 September 1976) where surface thermodynamic temperatures were $>15°C$ and there was no synoptic-scale rainfall. Almost all pixels were classified by the algorithm as rain over land. An examination of the ESMR 6 vertically polarized T_B's showed that the temperatures were below 0°C. Since calibration of the nighttime data is better than that of daytime data (Wilheit, 1978), this anomaly may be attributed to the changes in the surface emissivity caused by the presence of dew on the vegetation. The 0600 GMT National Weather Service map indicated that the conditions were ideal for the formation of dew. A large anticyclone centered over Virginia produced clear skies, winds less than 5 kt and dew

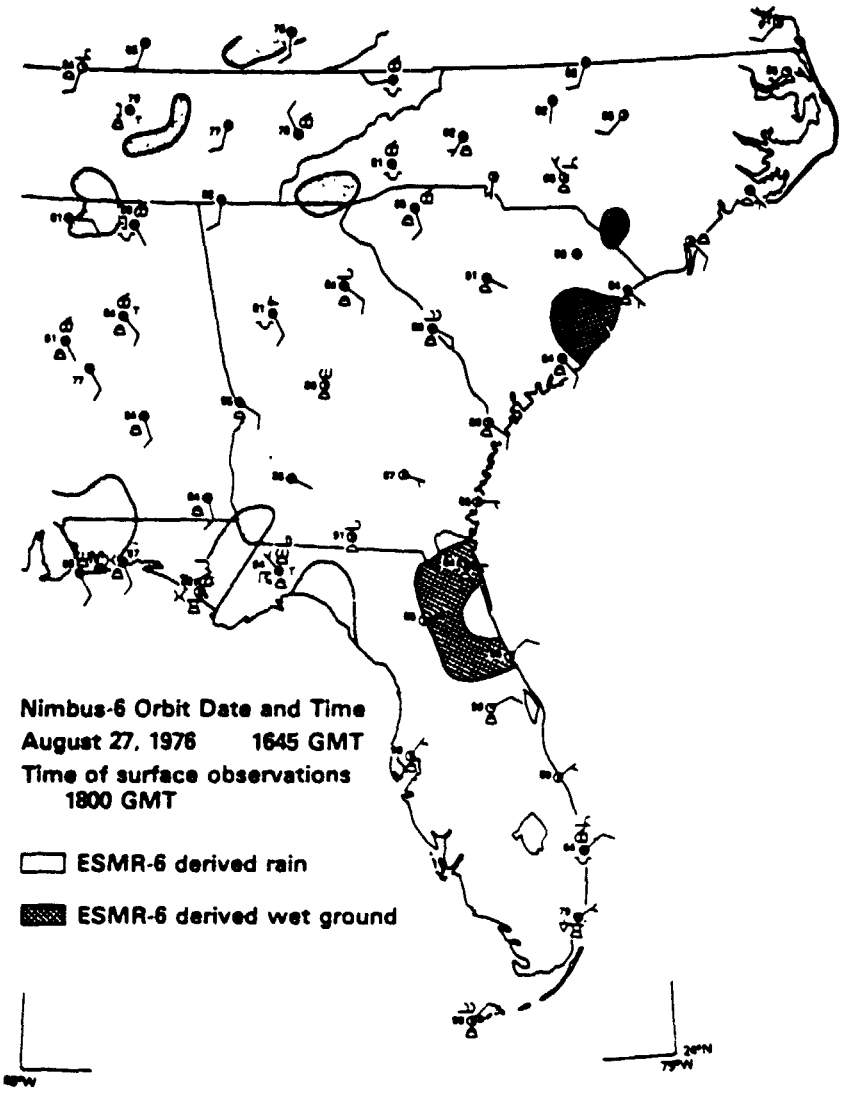


FIG. 10. As in Fig. 9 except at 1645 GMT 27 August 1976

point temperature differences of less than 3°C over the majority of the reporting stations in the southeast United States. Therefore, the classification algorithm trained by data sampled from Nimbus 6 daytime passes can be employed only when dew is absent.

8. Conclusion

Statistical analyses were performed on the sampled ESMR 6 data for the purpose of delineating rainfall areas over land from dry and wet land surfaces. It was found that synoptic-scale rainfall over land, where surface thermodynamic temperatures were greater than 15°C and where the vegetation was not covered with dew, could be delineated despite the large ESMR 6 IFOV. However, there was some ambiguity in distinguishing between rainfall over land and wet land surfaces.

REFERENCES

Barrett, E. C., 1970: The estimation of monthly rainfall from satellite data. *Mon. Wea. Rev.*, 98, 322–327.

——, 1973: Forecasting daily rainfall from satellite data. *Mon. Wea. Rev.*, 101, 215–222.

Bennett, B. M., 1951: Note on a solution of the generalized Behrens-Fisher problem. *Ann. Inst. Statist. Math.*, 2, 87–90.

Bond, A. D., and R. J. Atkinson, 1972: An integrated feature selection and supervised learning scheme for fast computer classification of multispectral data. *Remote Sensing of Earth Resources*, Vol. 1, F. Shahrokhi, Ed., The University of Tennessee Press, 645–672.

Born, M., and E. Wolf, 1975: *Principles of Optics*. Pergamon Press, 182 pp.

Cochran, W. G., 1952: The χ^2 test of goodness of fit. *Ann. Math. Statist.*, 23, 315–345.

Duda, R. O., and P. E. Hart, 1973: *Pattern Classification and Scene Analysis*. Wiley, 482 pp.

Fisher, R. A., 1938: The statistical utilization of multiple measurements. *Ann. Eugenics*, 8, 376–386.

Follansbee, W. A., and V. J. Oliver, 1975: A comparison of infrared imagery and video pictures in the estimation of daily rainfall from satellite data. NOAA Tech. Memo. NESS 62, 14 pp.

Fu, K. S., D. A. Landgrebe and T. L. Phillips, 1969: Information processing of remotely sensed agricultural data. *Proc. IEEE*, 57, 631–653.

Griffith, G. C., W. L. Woodley, S. Browner, J. Teijeiro, M. Maier, D. W. Martin and D. N. Sikdar, 1978: Rain estimation from geosynchronous satellite imagery—visible and infrared studies. *Mon. Wea. Rev.*, 106, 1153–1171.

Gunn, K. L. S., and T. U. R. East, 1954: The microwave properties of precipitation particles. *Quart. J. Roy. Meteor. Soc.*, 80, 522–545.

Hall, C. D., R. Davies and J. A. Weinman, 1978: The distribution of precipitation derived from Nimbus-6 data. Paper presented at 18th Conf. Radar Meteorology, Atlanta, Amer. Meteor. Soc.

Idso, S. B., R. D. Jackson and R. J. Reginato, 1975: Detection of soil moisture by remote surveillance. *Amer. Sci.*, 63, 549–557.

Kshirsagar, A. M., 1972: *Multivariate Analysis*. Marcel Dekker, 534 pp.

Martin, D. W., and W. D. Scherer, 1973: Review of satellite rainfall estimation methods. *Bull. Amer. Meteor. Soc.*, 54, 661-674.

Martin, D. W., J. Stout and D. N. Sikdar, 1975: GATE area rainfall estimation from satellite images. Rep. NOAA Grant 04-5-158-47, University of Wisconsin, 28 pp.

McFarland, M. J., and B. J. Blanchard, 1977: Temporal correlation of antecedent precipitation with Nimbus-5 ESMR brightness temperature. *Preprints 2nd Conf. Hydrometeorology*, Toronto, Amer. Meteor. Soc., 311–315.

Meeks, M. L., and A. Lilley, 1963: The microwave spectrum of oxygen in the earth's atmosphere. *J. Geophys. Res.*, 68, 1683–1703.

Meneely, J. M., 1975: Application of the Nimbus-5 ESMR to rainfall detection over land surfaces. Contract Rep. NAS 5-20878, 48 pp.

Mie, G., 1908: Beiträge zur Optik trübes media, Speziell kolloidaler Metallösungen. *Ann. Phys.*, 26, 597–614.

Okamoto, M., 1963: An asymptotic expansion for the distribution of the linear discriminant function. *Ann. Math. Statist.*, 34, 1286–1301.

Savage, R. C., and J. A. Weinman, 1975: Preliminary calculations of the upwelling radiance from rain clouds at 37.0 and 19.35 GHz. *Bull. Amer. Meteor. Soc.*, 56, 1272–1274.

——, P. J. Guetter and J. A. Weinman, 1976: The observation of rain clouds over land in Nimbus-6 electrically scanned microwave radiometer (ESMR-6) data. *Preprints 7th Conf. Aerospace and Aeronautical Meteorology and Symp. on Remote Sensing from Satellite*, Melbourne, Amer. Meteor. Soc., 131–136.

Scheffé, H., 1959: *The Analysis of Variance*. Wiley, 477 pp.

Scofield, R. A., and V. J. Oliver, 1977: A scheme for estimating convective rainfall from satellite imagery. NOAA Tech. Memo. NESS 86, 47 pp.

Shenk, W. E., R. J. Holub and R. A. Neff, 1976: A multispectral cloud type identification method developed for tropical ocean areas with Nimbus-3 MRIR measurements. *Mon. Wea. Rev.*, 104, 284–291.

Staelin, D. A., 1966: Measurements and interpretation of the microwave spectrum of the terrestrial atmosphere near one centimeter wavelength. *J. Geophys. Res.*, 71, 2875–2881.

Weinman, J. A., and P. J. Guetter, 1977: Determination of rainfall distribution from microwave radiation measured by the Nimbus-6 ESMR. *J. Appl. Meteor.*, 16, 437–442.

Wilheit, T. T., 1975: The electrically scanning microwave radiometer (ESMR) experiment. *Nimbus-6 Users Guide*, NASA Goddard Space Flight Center, 87–108.

——, 1978: The effect of wind on the microwave emission from the ocean's surface at 37 GHz. NASA/GSFC Tech. Memo. 79588, 24 pp. (To be published in *J. Geophys. Res.*)

——, A. T. C. Chang, M. S. V. Rao, E. B. Rodgers and J. S. Theon, 1977: A satellite technique for quantitatively mapping rainfall rates over the oceans. *J. Appl. Meteor.*, 16, 551–560.

Paper 53

Reprinted from MONTHLY WEATHER REVIEW, Vol. 107, No. 5, May 1979
American Meteorological Society
Printed in U. S. A.

The Benefits of Using Short-Interval Satellite Images to Derive Winds for Tropical Cyclones

EDWARD RODGERS

NASA/GSFC, Greenbelt, MD 20771

R. CECIL GENTRY

General Electric Space Division (MATSCO), Beltsville, MD 20705

WILLIAM SHENK

NASA/GSFC, Greenbelt, MD 20771

VINCENT OLIVER

NOAA/NESS, Camp Springs, MD 20233

(Manuscript received 25 October 1978, in final form 5 February 1979)

ABSTRACT

During the 1975, 1976 and 1977 North Atlantic hurricane seasons, NOAA's National Environmental Satellite Service (NESS) and NASA's Goddard Space Flight Center (GSFC) conducted a cooperative program to determine the best resolution and frequency now available from satellite images for deriving winds to study and forecast tropical cyclones. Rapid-scan images were obtained in 1975 at 7.5 min interval from SMS 2 for Hurricane Eloise on 22 September and of tropical cyclone Caroline on 28, 29 and 30 August; in 1976 at 3 min intervals from GOES 1 for tropical storms Belle on 5 August and Holly on 25 October; and in 1977 at 3 min intervals from GOES 1 for tropical cyclone Anita on 30 and 31 August and 1 September. Cloud motions were derived from these images using the Atmospheric and Oceanographic Information Processing System (AOIPS) at GSFC. Winds that were derived from the movement of upper (~200 mb) and lower tropospheric (~900 mb) level clouds using rapid scan data were compared with the 15 and 30 min interval data. This was done using visible images having 1, 2, 4 and 8 km resolution for the areas within 650 km of the storm center for the 1975 and 1976 tropical cyclones. Greater than 10 (5) times as many clouds could be tracked to obtain winds at both levels using 3 and 7.5 min rapid-scan images as when using 30 min (15 min) interval images. In addition, by using the frequent images, it was possible to track a few bright areas within the central dense overcast which appeared to be moving with the winds at low levels. For Hurricanes Eloise and Caroline the winds that were derived by tracking these bright areas within the central dense overcast had speeds differing in the mean by only 2.5 m s^{-1} from the wind speed measured by aircraft flying at ~0.5 km above the surface in the same quadrant 4 h later. Full-resolution visible images (1 km) were needed to track slow moving low-level cloud elements, since on a degraded resolution image, subpixel movement would introduce additive inaccuracies to the wind measurements. Rapid-scan full-resolution GOES 1 data for tropical cyclone Anita (1977) provided representative wind fields only outside the central dense overcast at the lower tropospheric level. For this area aircraft-measured wind speeds differed in the mean again by only 2.5 m s^{-1}.

1. Introduction

The destruction caused by tropical cyclones affecting the United States has increased at an alarming rate. In the decade from 1965–74, the average annual cost from tropical cyclone damage was nine times greater, even after adjustments for inflation, than during the decade from 1915–24 (Gentry, 1966). By contrast, the loss of life caused by tropical cyclones decreased considerably after 1935 with the improvement of the warning service.

It is a common belief that improving the warning services requires improved numerical-dynamical models for forecasting hurricane motion and intensity and a better knowledge of the initial meteorological parameters. In particular, initial wind data at several levels from the surface to the lower stratosphere extending outward from the center to include the environment of the tropical cyclone are needed. Elsberry (1977) stated that the poor results in forecasting the translation speed from the coarse-grid version of his tropical cyclone prediction model

0027-0644/79/050575-10506.50

was attributable to the deficiency of wind data for the initial field. Elsberry also noted that additional wind data are not only needed around the tropical cyclone center but within the surrounding cloud system out to ~600 km from the center.[1] In addition, Hovermale and Livezey (1977) showed that the introduction of a more realistic outer circulation (300–1100 km from the center) to their modeled spinup storm would reduce their vector position errors. For the 1976 Atlantic tropical cyclone season, the vector position errors for the 36 and 48 h forecasts increased by approximately a factor of 3 for storms over data-void ocean areas as compared to storms nearer to coastal stations.

Data outside 600 km from the center are usually obtained from radiosonde networks, transoceanic aircraft flights and from cloud motion wind analyses at approximately the 900 and 200 mb levels obtained from operational scanning geosynchronous satellites. Within the tropical cyclone cloud systems the only real source of wind data is the hurricane reconnaissance aircraft flight. However, the quantity of wind data obtained from these flights is not sufficient for these numerical-dynamical models and, in fact, the tendency in recent years is to reduce the quantity by decreasing the number of flights. Therefore, there is a great need to obtain additional wind data from another source.

An excellent source for the upper troposphere has been the motion of clouds tracked with successive infrared and visible images obtained operationaliy from the family of geosynchronous satellites at ~30 min intervals. However, less success has been found using cloud tracking to obtain winds at lower levels near tropical storms. One reason for the lack of success is that many of the clouds of the type and size best suited for tracers at low levels do not persist or maintain their identity for even 30 min. This has been observed by many previous investigators. For example, Fujita *et al.* (1975) observed that cloud turrets ranging between 0.5 to 3.2 km in size which they found to be the best targets to infer winds within the subcloud layer were hard to track targets in a tracking point of view because they were relatively short lived. Another problem in tracking low-level clouds within the area of the tropical cyclone cloud system is that these low-level clouds that are sometimes seen through the breaks in the cirrus frequently are only observable for a short period. Therefore, in order to supply the numerical-dynamical models with additional cloud-derived low-level winds within the tropical cyclone cloud system, more frequent observations from the geosynchronous satellite should be obtained.

To verify this hypothesis, the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA) initiated a cooperative project to use rapid-scan satellite imagery to obtain more wind data in tropical cyclones during the 1975–77 hurricane seasons. The goals of the project were to obtain more winds needed by the hurricane researchers (including models) and forecasters and to determine the optimum space and temporal resolution of the geosynchronous satellite data used to derived winds. Limited-scan visible and infrared images were obtained at 7.5 min intervals from the Synchronous Meteorological Satellite (SMS 2) in 1975 and at 3 min intervals from the Geostationary Operational Environmental Satellite (GOES 1) during the 1976 and 1977 tropical cyclone season for a few hours on selected tropical cyclone days. These were for tropical cyclones Eloise (22 September 1975), Caroline (28, 29 and 30 August 1975), Belle (5 August 1976), Holly (26 October 1976) and Anita (30, 31 August and 1 September 1977). Winds were obtained at the lower and upper tropospheric levels.

2. Analysis

The hypothesis that cumulus and cirrus cloud translation approximates the speed and direction of the ambient flow at the cloud-base level has been tested using aircraft observations in the trade wind regime over ocean areas (Hasler *et al.* 1977). Movement of cumulus clouds 3–15 km in diameter with bases at 960 mb and tops at 600–700 mb had a vector difference from the ambient flow at cloud base of 1.3 m s^{-1}. The average difference found between a limited number of cirrus clouds and the ambient flow within the cloud layer was 1.6 m s^{-1}.

The clouds for the five tropical cyclones were tracked on NASA's Atmospheric and Oceanographic Information Processing System (AOIPS) using a series of infrared or visible SMS 2/GOES 1 satellite digitized images (Billingsley, 1976). This system allows a user to interactively modify digital images and to display the results on a television monitor through utilization of a computer software package called the Meteorological Data Processing Package (METPAK). To track clouds using the METPAK software, a sequence of images at a given time interval must first be registered by aligning recognizable land features in each image. A correction is made for image distortion caused by satellite oscillating orbit motion and the axes not being parallel to the earth's orbit. The recognizable land feature is earth located by utilizing a navigation algorithm adapted from the University of Wisconsin's McIDAS navigation routine (Smith, 1975). The algorithm translates image coordinates (pixel, lines) to earth coordinates (latitude, longitude) with respect to the land feature. The images are then displayed in time lapse sequence on the television

[1] Personal communication, 1977.

monitor. Either the centroid or an identifiable point on the cloud that can be seen on all images (at least three images are used) are tracked subjectively by a moving cursor or objectively using an image correlation mode. The displacement of the cloud in earth coordinates divided by the time interval between images gives the wind velocity.

Other options in the METPAK software package involve variations of the spatial resolution of the image and determination of cloud height. The cloud height is calculated by estimating the cloud optical thickness for the observed cloud brightness and the infrared emissivity using Kirchoff's law. The cloud optical thickness is calculated by a multiple-scattering program designed for the McIDAS at the University of Wisconsin (Smith, 1975). The emissivity and the percent of cloud coverage as obtained from the visible image are used to determine the cloud top temperature. The cloud top is then estimated using the standard atmosphere corrected for latitude and date. The cloud height was used to assign the level for which the clouds were tracked. Clouds above the 350 mb level were assigned to the upper troposphere, while those below 700 mb were accepted as low tropospheric clouds except near the center of tropical cyclones. The few clouds found between these levels were not traced.

The spatial and temporal resolutions that were used for clouds tracked for both the lower and upper troposphere for each of the five tropical cyclones are listed in Table 1. Time of observation during the daylight hours was approximately between 1300 to 1900 GMT. Visible images (designated by V in table) were primarily used to track clouds

TABLE 1. Spatial and temporal resolution of the SMS 2/GOES 1 imagery used to track clouds for each of the five tropical cyclones

Space resolution (km)	Temporal resolution (min)		
	3–5	15	30
	Low level		
4	Eloise (22/9/75) V		
2	Eloise (22/9/75) V Caroline (30/8/75) V Caroline (28/8/75) V	Eloise (22/9/75) V	Eloise (22/9/75) V Caroline (30/8/75) V
1	Eloise (22/9/75) V Caroline (30/8/75) V Caroline (28/8/75) V Belle (5/8/76) V Holly (26/10/76) V Anita (30–31/8/77) V	Eloise (22/9/75) V Caroline (30/8/75) V	Eloise (22/9/75) V Caroline (30/8/75) V Caroline (28/8/75) V Belle (5/8/75) V Holly (26/10/76) V
	High level		
4	Eloise (22/9/75) IR		Eloise (22/9/75) IR
2	Eloise (22/9/75) V Caroline (30/8/75) V Caroline (28/8/75) V Belle (5/8/76) V		Eloise (22/9/75) V Caroline (30/8/75) V Caroline (28/8/75) V Belle (5/8/76) V

TABLE 2. Hurricane Eloise 22 September 1975. Number of low-level wind vectors.

	Distance from center (km)		
	0–222	0–444	0–666
30 min 2 km V	0	1	
1 km V	0	1	9
15 min 2 km V	0	5	
1 km V	0	6	19
7.5 min 4 km V	4	34	
2 km V	4	38	
1 km V	4	58	119

at both levels; however, the infrared (designated by IR in the table) was used to track upper tropospheric clouds for tropical cyclone Eloise in a simulation of nighttime coverage.

3. Results

The results from the SMS 2 images of tropical cyclones Eloise and Caroline are presented in Tables 2–5. Low tropospheric clouds were tracked for both tropical cyclones using at least three successive visible images with spatial resolutions of 1, 2 and 4 km. The time intervals between images were 7.5, 15 and 30 min. No cloud element was tracked that could not be delineated in all images of each sequence.

Table 2 shows the number of low tropospheric clouds tracked for Hurricane Eloise at radial distances of 222, 444 and 666 km from the center for a given spatial and temporal resolution. Improved temporal resolution substantially increased the number of traceable low tropospheric cloud elements. The shorter interval data not only improved the cloud temporal continuity but also eliminated the ambiguity caused by tracking cloud growth rather than displacement. From the center out to 666 km, the 7.5 min interval images increased the number of traceable cloud elements by more than a factor of 10 (5) over that of the 30 min (15 min) interval images. Within 222 km from the center, cloud elements could only be tracked using shorter interval images.

Figs. 1, 3 and 4 illustrate the distribution and the number of wind vectors using 1 km spatial resolution images at time intervals of 7.5, 15, and 30 min, respectively. Improved temporal resolutions not only increased the number of traceable cloud elements but also corrected an erroneous wind [west-northwest wind 444 km southeast of the center (see Fig. 4)] caused by tracing cloud growth rather than cloud motion and added wind information within the eye wall.

Improved spatial resolution in the case of Eloise did benefit the tracking of small stratocumulus elements in the stable air west of the center and

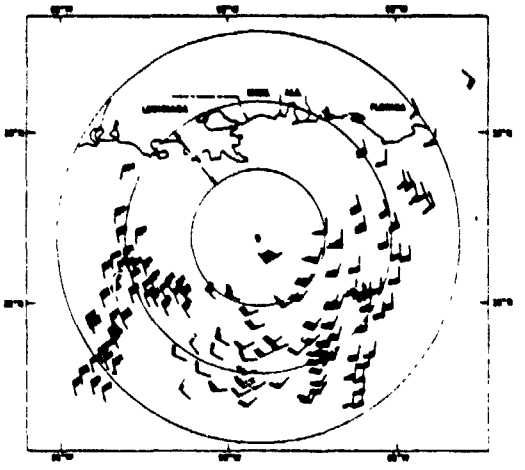


FIG. 1. Low-level winds (kt) derived from cloud movements between three successive SMS 2 pictures with 1 km resolution and 7.5 min intervals (1842, 1850, 1857 GMT) on 22 September 1975 of Hurricane Eloise

small cumulus elements south and east of the center. At spatial resolution >1 km, no attempt was made to track clouds outside the radius of 444 km from the center of Hurricane Eloise. Full-resolution (1 km) visible images were also found to be important for tracking low tropospheric clouds within the central dense overcast where the clouds are difficult to delineate because of the cirrus; and over the land where the sometimes slower moving clouds moved less than one television display pixel using the 2 and 4 km resolution. In using the infrared images, little information was obtained for tracing lower clouds even at full spatial resolution. Figs. 1 and 2 illustrate the distribution and the number of wind vectors that are derived from the motion of low-level cloud elements at 1 and 4 km spatial resolution using visible images at 7.5 min intervals.

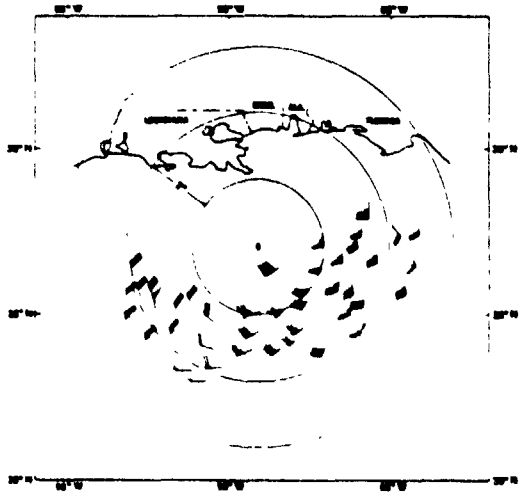


FIG. 2. As in Fig. 1 except with 4 km resolution.

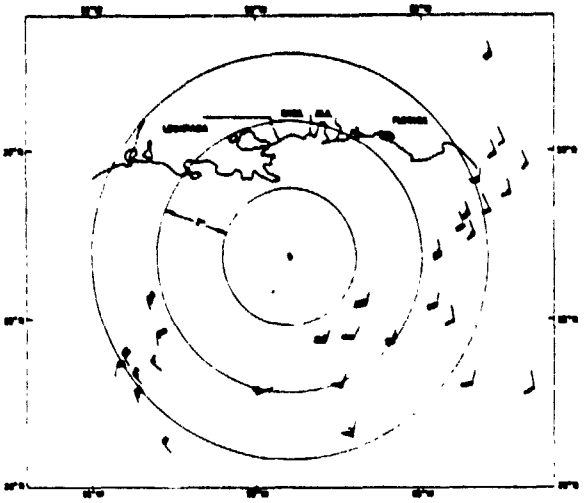


FIG. 3. As in Fig. 1 except with 1 km resolution and 15 min intervals (1842, 1857, 1912 GMT).

Cirrus clouds were examined for Hurricane Eloise using visible and infrared images at 7.5 and 30 min intervals. The purpose of using the infrared images was to determine if rapid scan images could be of any greater value at night than the 30 min interval images. The spatial resolution used for the visible and infrared images were 2 and 8 km, respectively. No additional information was obtained when tracking upper tropospheric clouds by using visible images at 1 km as compared to 2 km resolution.

Table 3 summarizes the results for the high clouds. Within 666 km from the center, approximately twice as many clouds could be tracked using the 2 km resolution visible as compared to the 8 km resolution infrared imagery. Thus, the structure of cirrus clouds is best delineated with higher spatial

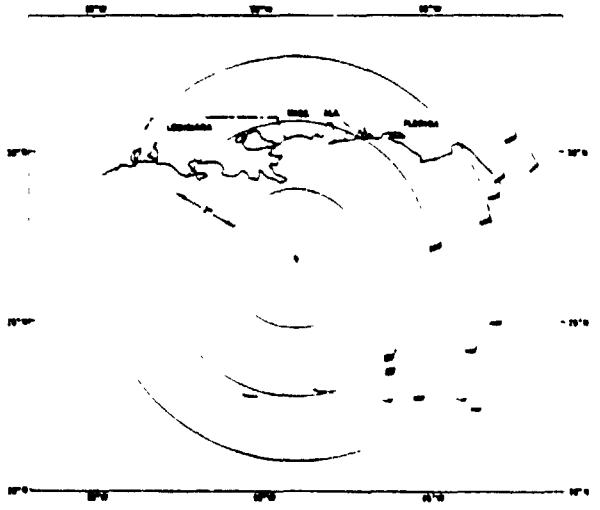


FIG. 4. As in Fig. 1 except with 1 km resolution and 30 min intervals (1820, 1850, 1920 GMT)

TABLE 3. Hurricane Eloise 22 September 1975 1900 GMT. Number of high-level wind vectors.

	Distance from center (km)		
	0–222	0–444	0–666
30 min 8 km IR	0	3	6
2 km V	0	5	9
7.5 min 8 km IR	1	13	30
2 km V	10	44	71

resolution visible images. As was found with the low tropospheric clouds, the improved temporal resolution again increased the number of traceable cloud elements considerably (by a factor of 5 in this case). Figs. 5–8 illustrate the distribution and number of winds that were derived from tracking high clouds using visible and infrared images at 7.5 and 30 min intervals. The table and figures show that using low-resolution infrared images at short intervals greatly enhances the number of traceable cloud elements; thus infrared images would be more valuable for nighttime use if they were obtained more frequently.

One of the problems that was encountered in using longer interval visible and infrared images to track cirrus clouds was the ambiguity introduced by tracking cloud patterns that were repetitive. For example, in Fig. 9, an infrared image of Hurricane Eloise at 2029 GMT 22 September 1975, there is a repetitive saw-tooth pattern in the western portion of the cloud mass (seen adjacent to A in the figure). This entire pattern was observed moving northward when short-interval imagery was shown in a time-lapse mode. When the imagery was spaced at longer periods (15 and 30 min), the direction of the displacement and the motion was uncertain.

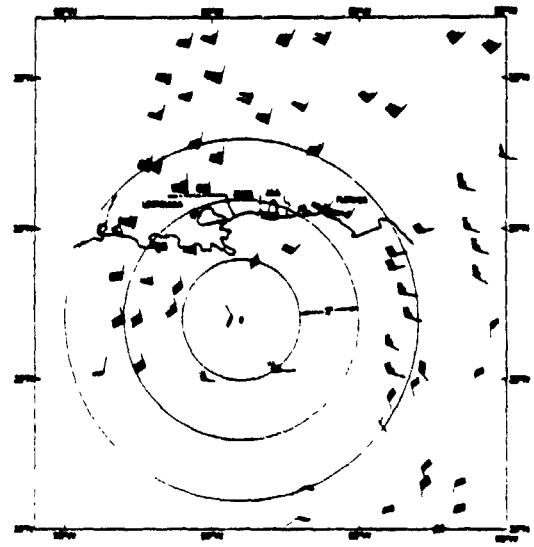


FIG. 6. As in Fig. 5 except with 8 km infrared resolution.

Tropical Cyclone Caroline was examined on all three days using visible images. Tables 4 and 5 describe the results for tracking lower and upper tropospheric clouds for Hurricane Caroline on 30 August 1975. Results for 28 and 29 August were quite similar. For the lower tropospheric clouds obtained from the 7.5 min interval visible images with 444 km radius from the center, the degradation in the spatial resolution from 1 to 2 km did not decrease the number of wind vectors because only the larger cloud elements were traceable. However, at radii >444 km, improved resolution did increase

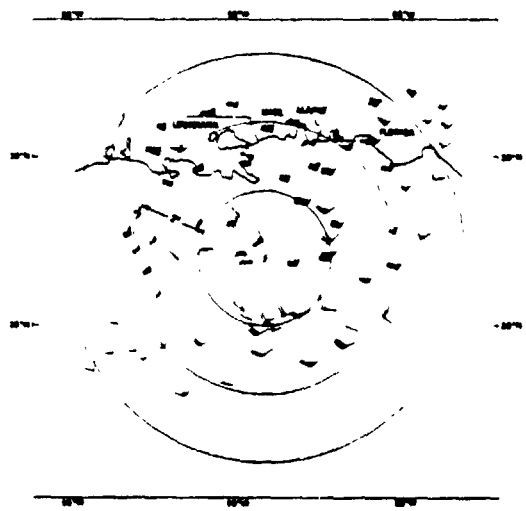


FIG. 5. High-level winds (kt) derived from cloud movements between three successive SMS 2 pictures with 2 km visible resolution and 7.5 min intervals (1842, 1850, 1857 GMT) on 22 September 1975 of Hurricane Eloise.

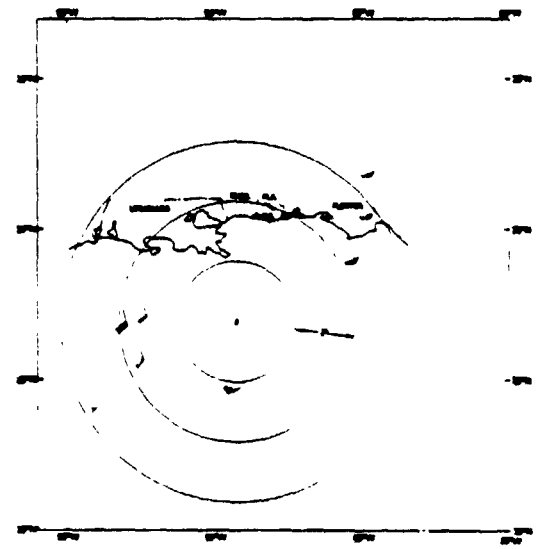


FIG. 7. As in Fig. 5 except with 2 km visible resolution and 30 min intervals (1820, 1850, 1920 GMT).

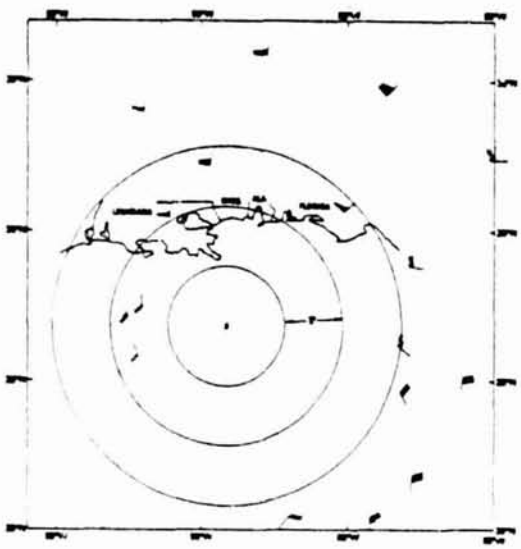


FIG. 8. As in Fig. 5 except with 8 km infrared resolution and 30 min intervals (1820, 1850, 1920 GMT).

the number of traceable cloud elements. This was because the clouds to the west were over land and moved less than a television display pixel per time period in the 2 km resolution images. Using visible images whose intervals were 30 min, low-level clouds over land could not be tracked because of the lack of temporal continuity. Improved temporal resolution again substantially increased the number of traceable cloud elements by a factor of 6 (<2) for 7.5 min interval images as compared with 30 min (15 min) interval images for an area whose radius was 666 km from the center. A factor of 3 improvement was found using 7.5 min interval visible images over that of 30 min for tracking upper tropospheric clouds. Because of the smoother texture of the cirrus canopy and therefore a lack of traceable elements, this improvement was not as dramatic as in the Eloise case.

FIG. 9. Infrared imagery of Hurricane Eloise at 1912 GMT 22 September 1975 (note serrated pattern at A).

TABLE 4. Hurricane Caroline 30 August 1975 1800 GMT. Number of low-level wind vectors.

	Distance from center (km)		
	0–222	0–444	0–666
30 min 2 km V	0	5	13
1 km V	0	3	14
15 min 1 km V	17	46	75
7.5 min 2 km V	23	65	88
1 km V	23	61	121

The results from tropical cyclones Belle and Holly obtained from the GOES 1 images during the 1976 hurricane season are presented in Table 6. The same format was followed for tracking clouds for these storms as was used for the tropical cyclones in 1975. The only difference was that the temporal resolution was compared only between 30 and 3 min intervals using visible images. No differences in spatial resolution was considered. It can be seen from Table 6 that the rapid scan images again dramatically increased (factor of 6) the number of wind vectors obtained from cloud motion in the upper and lower troposphere for Belle and lower troposphere for Holly when the image interval was reduced from 30 to 3 min. Tropical Storm Holly at this time had no high tropospheric clouds except in the southeast quadrant.

Figs. 10 and 11 dramatize the results for tropical storm Holly. The figures show the distribution and number of low tropospheric winds that can be derived from cloud motion using 30 min as compared to 3 min interval images. Because of the latitude belt for which the satellite was programmed to take scans, the images and therefore wind vectors in the

TABLE 5. Hurricane Caroline 30 August 1975 1800 GMT. Number of high-level wind vectors.

	Distance from center (km)		
	0–222	0–444	0–666
30 min 2 km V	0	14	22
7.5 min 2 km V	0	55	70

TABLE 6. Tropical cyclone Belle 5 August 1976 1800 GMT and tropical cyclone Holly 26 October 1976 1300 GMT. Number of low-level wind vectors.

	Distance from center (km)		
	0–222	0–444	0–666
	Low level		
Belle 30 min 1 km V	0	23	—
3 min 1 km V	8	110	—
Holly 30 min 1 km V	4	29	48
3 min 1 km V	80	225	285
	High level		
Belle 30 min 1 km V	0	18	—
3 min 1 km V	16	75	—

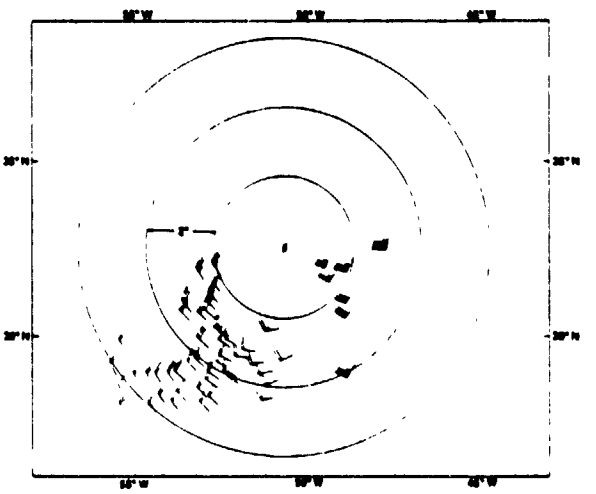


FIG. 10. Low-level winds (kt) derived from cloud movements between three successive GOES 1 pictures with 1 km resolution and 30 min intervals (1530, 1600, 1630 GMT) on 26 October 1976 of tropical storm Holly.

northern half of Holly were not available. Fig. 11 delineates an area of maximum winds (50 kt) to the east of the center obtained from rapid scan images which verify the tropical storm status reported by reconnaissance aircraft. This wind maximum was not observed from the 30 min interval data.

Rapid scan images are seen from this observation of tropical storm Holly to be particularly valuable for estimating the lower tropospheric wind field nearer to the center in weaker or developing tropical cyclones whose upper tropospheric clouds are not dense enough to obscure the lower level circulation. Even for stronger tropical cyclones that have a more developed high-level cloud shield the lower tropospheric wind field sometimes can be estimated near the center. This was true for tropical cyclones Eloise and Caroline. It is hypothesized that some deep convection near the storm center with small lateral area translates with the wind speed at the cloud-base level. A few of these tracers appearing as bright spots in the upper cirrus overcast can be tracked using high resolution (both space and temporal) imagery. Maritime trade cumulus clouds (tops as high as 200 mb) have been found from aircraft observation to move

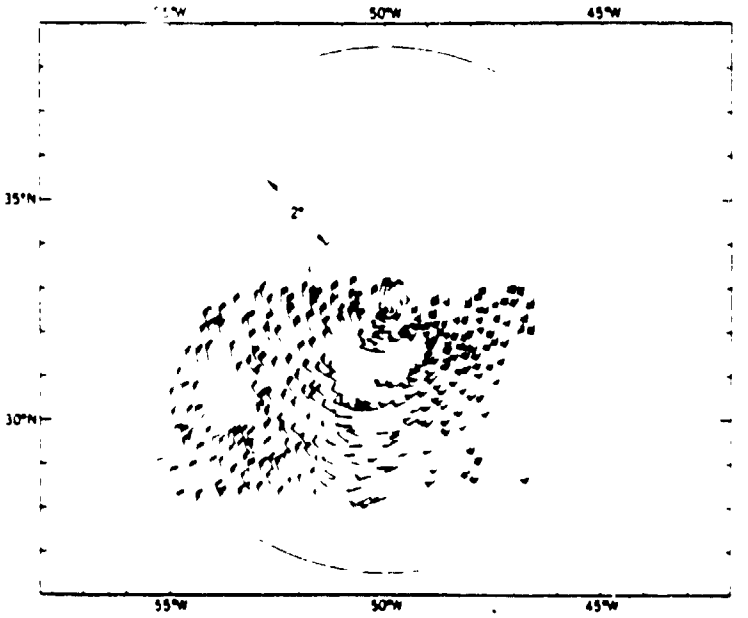


FIG. 11. Low-level winds (kt) derived from cloud movements between four successive GOES 1 pictures with 1 km resolution and 3 min intervals (1342, 1345, 1348, 1351 GMT) on 26 October 1976 of tropical storm Holly

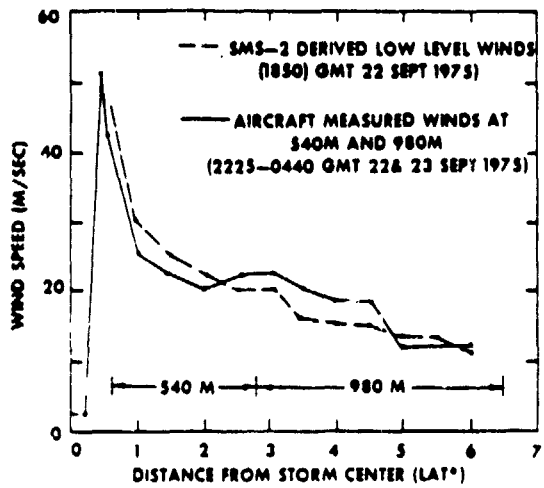


FIG. 12. Comparison of wind speeds derived from cloud motions between three consecutive SMS 2 satellite visible imagery at 7.5 min intervals with winds measured 4–6 h later by reconnaissance aircraft flying at 540 and 980 m in Hurricane Eloise on 22 September 1975.

very nearly with the ambient flow at the cloud base (Hasler *et al.* 1977). Although there is no aircraft verification that deep convective clouds move with the wind flow at cloud base within a tropical cyclone-type circulation, the small vertical shear and the large upward flux of angular momentum near the center suggest that it may happen (Frank, 1977). To examine this hypothesis, winds observed from the NOAA research aircraft at an approximate altitude of 1 km outside and 0.5 km inside the principal cloud shield for Hurricane Eloise (2245–0440 GMT 22 and 23 September 1975) and for Hurricane Caroline (1634–2332 GMT 30 August 1975) were compared with cloud-derived winds. Some small, bright convective cells that persisted during the time period were used to infer low-level winds near the center. However, the translation of many larger bright convective cells near the center that moved slower than the winds was suspected to be due to cloud growth rather than advection by the wind and were not used. The results that are seen in Figs. 12 and 13 show that the cloud-derived winds inside and outside the tropical cyclone cloud system compared reasonably well with aircraft-measured winds over the same area relative to the storm center even though the times differed by as much as 6 h. The cell tracked near the center of Eloise approximated Eloise's maximum winds [90 kt wind vector east-southeast of the center (Figs. 1 and 2)]. No cells suitable for tracking could be found within 80 km of Hurricane Caroline's center. The average difference in absolute speed of winds obtained by the two methods is approximately 2.5 m s^{-1} for each storm. Considering the time difference in the measurements, this result is no greater than would have occurred if the winds were obtained by aircraft for each case because of the great natural variability of winds in tropical cyclones (Gentry, 1964).

This good agreement between satellite and aircraft winds differ from the findings of Gentry *et al.* (1970) in which they tracked large bright convective cells near the center of Hurricane Gladys on 17 October 1968 using the 14 min interval visible images from the Multicolor Spin Scan Cloud Camera on ATS 3. It was found that these clouds moved from ⅓ to ½ of the speed of the low-level winds. The difference may be attributed to the ambiguity caused by cloud growth when tracking large cells using 4 km visible resolution images at a 14 min time interval.

Rapid-scan (3 min interval) full-resolution GOES 1 data for tropical cyclone Anita on 30 and 31 August and 1 September 1977 at approximately 1600 GMT were used to derive wind fields for both the cloud base and the upper troposphere. At the lower tropospheric level a representative wind field was obtained outside the central dense overcast. Within the central dense overcast, there were no low cloud elements that could be traced. Fig. 14 shows an image of tropical storm Anita for 1600 GMT 31 August 1977 together with the derived lower tropospheric wind field. The vector length represents cloud speed while the vector depicts the direction of cloud motion (the vector at the center of Anita locates the center of the storm). From this figure one can detect the directional convergence southwest of the eye. A speed maximum of 40 kt is west and south of the eye. Winds measured by the NOAA research aircraft were compared with the satellite-derived lower tropospheric winds where they could be matched in time and space.

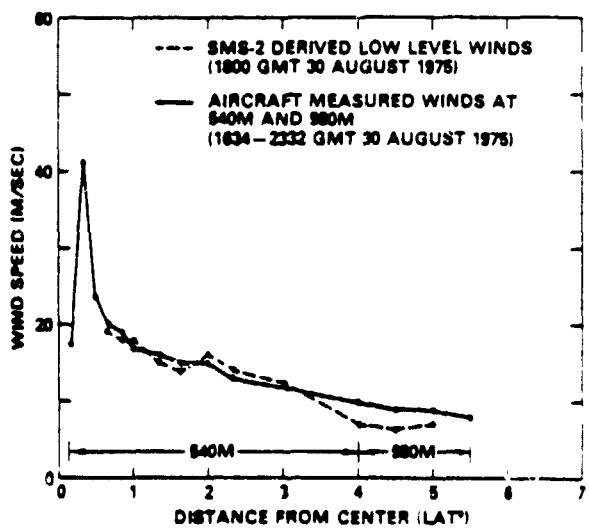


FIG. 13 Comparison of winds derived from cloud motion between three consecutive SMS 2 satellite 1 km visible imagery at 7.5 min intervals with winds measured 2–4 h later by reconnaissance aircraft flying at 540 and 980 m in Hurricane Caroline on 30 August 1975

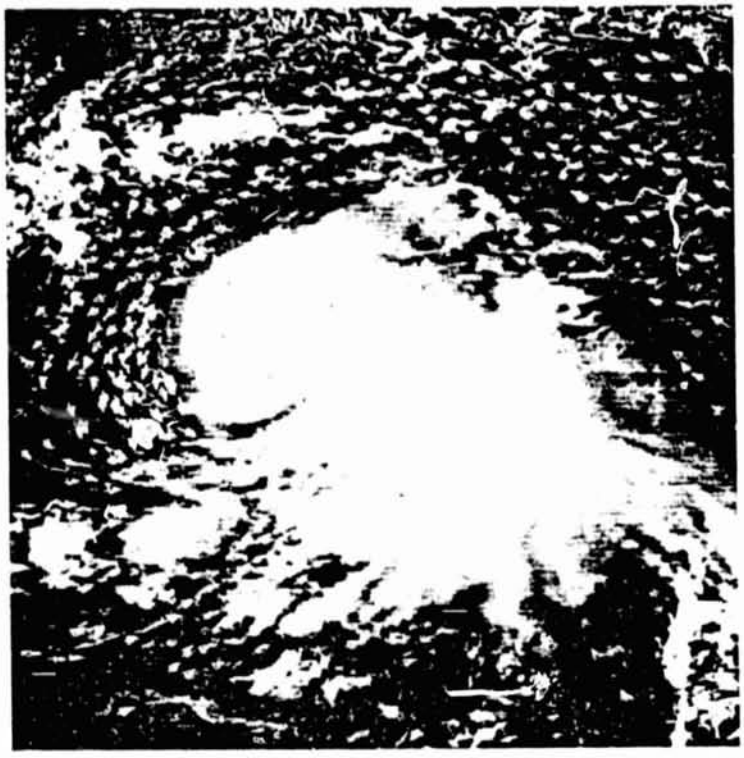


FIG. 14. Visible image of Tropical Storm Anita at 1600 GMT 31 August 1977 together with the rapid-scan (3 min interval) satellite-derived lower tropospheric wind field. Vector length represents cloud speed while vector direction represents the direction the cloud is moving. Vector at center locates the center of the storm.

This was on 31 August 1977 at approximately 1400 GMT when the aircraft flew southwest from New Orleans into the storm at an altitude of 0.5 km. Comparison was made between 220 and 440 km from the center in the northeast quadrant. Results are in Fig. 15 where again the average difference in absolute speed represented by the two methods is approximately 2.5 m s^{-1}. Although the agreement between cloud motion and wind is good in these cases, more events that have both rapid scan satellite and aircraft wind data are needed.

At the upper tropospheric level, (Fig. 16) limited scan data again aided in obtaining a representative wind field. Cirrus clouds could be used as tracers both outside and within the central dense overcast

4. Conclusion

High spatial and temporal resolution satellite imagery makes it feasible to provide a large number of lower and upper tropospheric winds which can be obtained by tracking clouds within 650 km of tropical cyclone centers. Up to 10 (5) times as many low-level winds were derived from images spaced at 3 or 7.5 min intervals as from those at 30 min (15 min) intervals. Greater space resolution is especially important when the low-level clouds are over land or imbedded in the dense high overcast near the storm center. As long as the clouds are over water and removed from the high-level clouds, 2 km space resolution is adequate. Rapid-scan

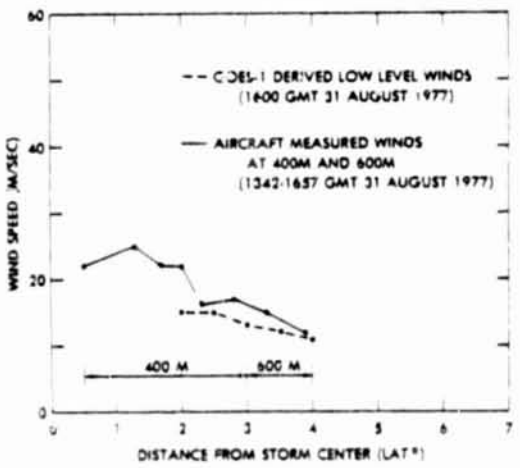


FIG. 15. Comparison of winds derived from cloud motion between three consecutive GOES 1 satellite visible imagery at 3 min intervals with winds measured 4–6 h later by reconnaissance aircraft flying at 540 and 980 m in tropical cyclone Anita on 31 August 1977.

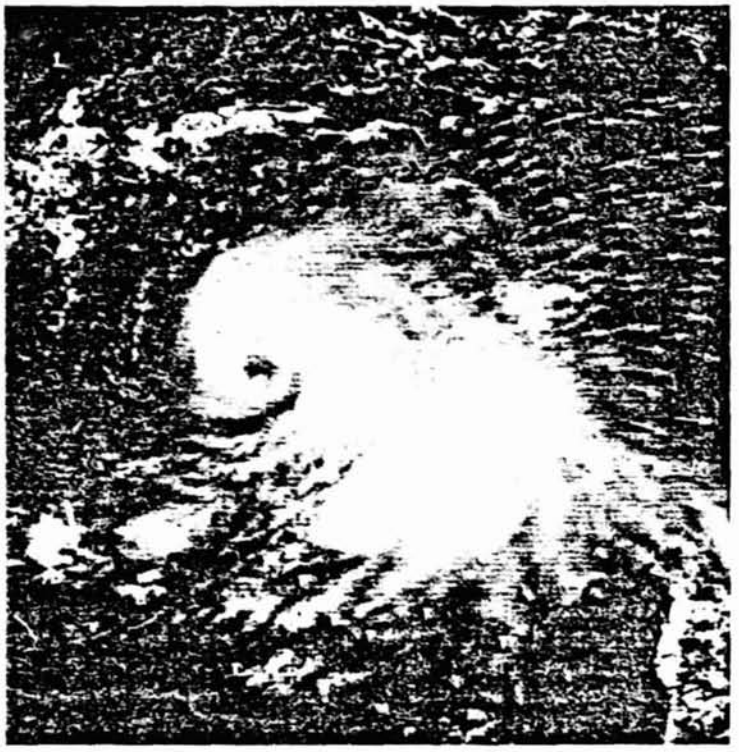


FIG. 16. As in Fig. 14 except vectors represent cloud motion and direction at the upper tropospheric level.

full-resolution infrared and visible images minimized the "erroneous winds" derived by tracking cloud elements that propagate by growing on one side and dissipating on the other and by tracking repetitive patterns that provided ambiguous indications of direction of movement.

The higher temporal resolution also made it possible to greatly increase the number of traceable upper tropospheric cloud elements both during the day and night. With the rapid-scan full-resolution visible images it was sometimes possible to derive low-level winds within the central dense overcast near the center of tropical cyclones.

Short-interval (7.5 min or less) full-resolution images (≤2 km) make it feasible to provide many of the wind data not available from other sources between 200 and 1100 km from tropical cyclone centers. These winds are urgently needed to initialize dynamical-numerical model predictions. This additional source of data could provide the input such that the numerical models will give improved forecasts.

Acknowledgment. The authors wish to thank Dr. Robert Sheets of NEHML for making available the wind data obtained the NOAA research aircraft for tropical cyclones Eloise and Caroline 1975 and Anita 1977.

REFERENCES

Billingsley, J. B., 1976: Interactive images for meteorological application at NASA Goddard Space Flight Center. *Preprints 7th Conf. Aerospace and Aeronautical Meteorology and Symposium on Remote Sensing from Satellites*, Melbourne, Amer. Meteor. Soc., 268–275.

Elsberry, R. L., 1977: Operational data tests with a tropical cyclone model. Navy Tech Rep. NPS-63Es77031, 1–30 [NTIS AD-A 039 183].

Frank, W. M., 1977: The structure and energetics of the tropical cyclone I. Storm structure. *Mon. Wea. Rev.*, **105**, 1119–1135.

Fujita, T. Theodore, E. W. Pearl and W. E. Shenk, 1975: Satellite-tracked cumulus velocities. *J. Appl. Meteor.*, **14**, 407–413.

Gentry, R. Cecil, 1964: A study of hurricane rainbands. National Hurricane Res. Project Rep. No. 69, U.S. Dept. of Commerce, 85 pp. [NTIS PB-168 417].

——, 1966: Nature and scope of hurricane damage. *Hurricane Symposium*, Amer. Soc. Oceanogr., Publ. No. 1, 229–254.

—— T. T. Fujita and R. Sheets, 1970: Aircraft, spacecraft, satellite and radar observation of Hurricane Gladys 1968. *J. Appl. Meteor.*, **9**, 837–850.

Hasler, A. F., W. E. Shenk and W. Skillman, 1977: Wind estimation from cloud motion: Phase 1, 2 and 3 of an *in situ* aircraft verification experiment. *J. Appl. Meteor.*, **18**, 812–815.

Hovermale, J. B., and R. E. Livezey, 1977: Three-year performance characteristics of the NMC hurricane model. *Preprints 11th Tech. Conf. Hurricanes and Tropical Meteorology*, Miami, Amer. Meteor. Soc., 122–125.

Smith, E. A., 1975: The McIDAS System. *IEEE Trans. Geo-Science Electron.*, **13**, 123–136.

Reprinted from *Journal of Applied Meteorology*, 17, pp. 458-476, 1978.

Nimbus 3/ATS 3 Observations of the Evolution of Hurricane Camille

WILLIAM E. SHENK AND EDWARD B. RODGERS

Goddard Space Flight Center, Greenbelt, Md. 20771

(Manuscript received 14 June 1977, in final form 28 November 1977)

ABSTRACT

Three periods within the life cycle of Hurricane Camille (1969) are examined with radiometric and camera measurements from Nimbus 3 and camera information from ATS 3 in conjunction with conventional information. These periods are the deepening phase, the interaction of Camille with mid-latitude westerlies, and the excessive rain-producing period when the cyclone was over the central Appalachians.

Just prior to significant deepening, the Nimbus 3 Medium Resolution Infrared Radiometer (MRIR) window and water vapor channels showed a band of developing convection that extended to the cirrus level in the southeastern quadrant of the storm which originated from the ITCZ. Low-level wind fields were derived from conventional sources as well as from cumulus clouds tracked from a series of ATS 3 images. Within this band were low-level 30 kt winds that supplied Camille with strong inflow where the air passed over sea surface temperatures that were 1-3 standard deviations above normal.

At the beginning of the rapid deepening the MRIR radiometer measurements indicated a rapid contraction of the central dense overcast and then an expansion as the maximum deepening rate occurred. Simultaneously, the increase in the MRIR equivalent blackbody temperatures (T_{BB}) indicated the development of large-scale subsidence throughout the troposphere northwest of the center. When Camille weakened as it moved over the lower Mississippi Valley, the cyclone acted as a partial obstruction to the synoptic-scale flow and increased the subsidence west and north of the cyclone center as indicated by the increase in water vapor T_{BB} and verified by three-dimensional trajectories. Increased cloud-top elevations, approaching the levels reached when Camille was an intense cyclone over the Gulf of Mexico, were estimated from the Nimbus 3 High Resolution Infrared Radiometer (HRIR) measurements on 20 August 1969, when Camille produced rains of major flood proportions near the east slopes of the Appalachians in central Virginia.

1. Introduction

From a meteorological point of view, Hurricane Camille was the most important Atlantic hurricane of record (Simpson *et al.*, 1970). No recording anemometer equipment was able to survive in the area of maximum winds as the cyclone moved onshore. Wind velocities approaching 175 kt were estimated from an appraisal of the damage within a few hundred meters of the coast. The minimum central pressure of 905 mb was second only to the Florida Keys storm of 1935 (892 mb) for Atlantic hurricanes. Whereas the Florida Keys storm probably had higher maximum winds than Camille, the diameter of hurricane winds was less than 100 km, as compared to 160 to 240 km for Camille.

Camille began to develop near the island of Grand Cayman from a tropical disturbance which moved westward across the Atlantic from the African Coast (Fig. 1). The cyclone deepened rapidly as it passed over the western tip of Cuba into the Gulf, and it reached maximum intensity by 0000 GMT[1] 17 August 1969. This intensity was maintained with little apparent fluctuation as observed by reconnaissance aircraft until the center reached the coast at 0430 on 18 August. Within 12 h, the winds had decreased to less than hurricane strength as the cyclone moved northward through Mississippi. For the next 36 h the remains of Camille moved north, then east, with the winds and rains diminishing in a typical history of a tropical cyclone moving away from its moisture source. Suddenly, excessive rainfall began over the central Appalachians in response to an injection of extremely moist air at low levels, an unstable lapse rate, upper tropospheric horizontal divergence and the topography (Schwarz, 1971). This portion of Camille's history is similar to that of Hurricane Diane which in 1955 produced massing flooding in New England although she had weakened considerably after moving across the Carolinas and Virginia. Camille briefly regained tropical storm strength after moving eastward off the Atlantic Coast near Norfolk, but interaction with a cold front caused the cyclone to become extratropical.

Throughout the important phases of Camille's life cycle, the storm was observed by meteorological satellites such as Nimbus 3, ATS-3, and the operational satellites. This outstanding coverage, along with conventional measurements, permitted the examination in some detail of three important events in the life cycle of Camille: 1) the rapid deepening phase, 2) the interaction with the mid-latitude westerlies as the cyclone moved inland, and 3) the development of the heavy rains over the central Appalachians.

2. Available data and mapping procedures

Radiometric measurements with a 55 km spatial resolution (at nadir) were taken at 12 h intervals by

[1] All times GMT unless noted otherwise.

0021-8952/78/0458-0476$09.50

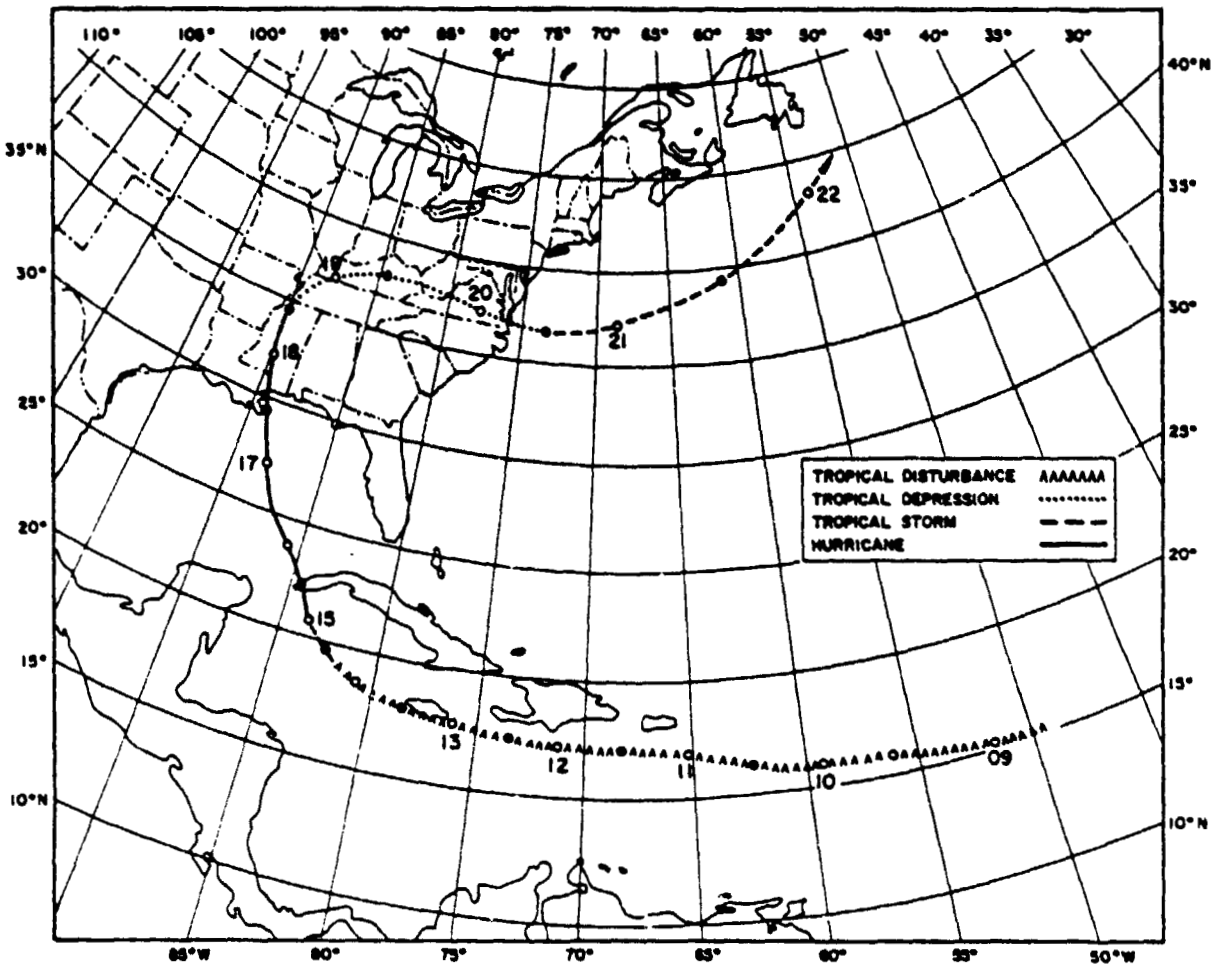


FIG. 1. Track and development stages of Hurricane Camille from 9–22 August 1969.
(Extracted from *Mon. Wea. Rev.*, 89, p. 294.)

the Nimbus 3 Medium Resolution Infrared Radiometer (MRIR) in five spectral regions. Four of these channels measured at 0.2–4.0 μm, 6.5–7.0 (6.7) μm, 10–11 (10.5) μm and 20–23 (21.5) μm, respectively. The first channel measured nearly all of the reflected solar energy, the third is in an infrared atmospheric window, and the second and fourth measure the emitted energy in regions of moderate to strong water vapor absorption. Fig. 2 shows the layers of a mean tropical, cloudless atmosphere that contribute to the sensed radiance in each of the three IR spectral intervals. The emission in the 6.7 μm region primarily comes from the upper troposphere, the 21.5 μm emission emanates mostly from the middle troposphere, and the 10.5 μm radiance nearly all comes from the surface and lower troposphere.

Nimbus 3 High Resolution Infrared Radiometer (HRIR) nighttime emission measurements were made at 3.5 to 4.1 μm with a 9 km subsatellite-track spatial resolution. Further details of the Nimbus 3 satellite and the instrumentation are provided in the *Nimbus 3 User's Guide* (Nimbus Project, 1969).

Visible measurements were obtained at 11 min intervals from the Multicolor Spin Scan Cloud Camera (MSSCC) on ATS 3. Due to the frequent observations, it was possible to track the motions of individual cloud elements to determine the wind. The clouds were classified as high or low, depending on their appearance and motion; the high clouds most likely indicating the air motion near 200 mb and the low clouds near 950 mb (Hasler *et al.*, 1976). Some middle clouds may have inadvertently been included, especially in the high-cloud sample. Hubert and Whitney (1971) have discussed cloud motion-wind accuracies from ATS data. More information on the ATS 3 satellite and the MSSCC can be found in the *ATS User's Guide* and catalog publication series.

The Nimbus radiation measurements were mapped in the stereographic horizon map projection (Shenk *et al.*, 1971) where the center of Camille was always placed at the center of the map. Thus, Camille was mapped (but not viewed) in the same perspective throughout the period of interest. A correction was applied to the measurements from the water vapor channels (6.7 and 21.5 μm) to compensate for limb-

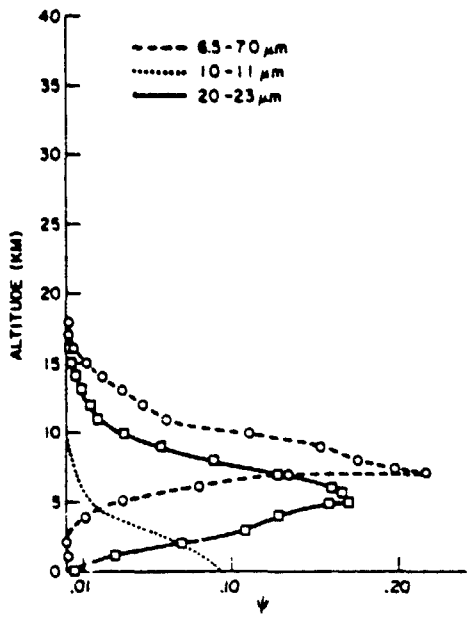


FIG. 2. Weighting function (ψ) curves [ψ units are W (cm² sr cm^{-1} km)$^{-1}$ normalized to unity] determined from radiative transfer theory based on a mean tropical atmosphere for three Nimbus MRIR channels (6.7 10.5 and 21.5 μm).

darkening effects.[1] The correction varied from zero to several degrees Kelvin for equivalent blackbody temperatures (T_{BB}) measured at a nadir angle of 50° (the maximum allowed).

Conventional radiosonde and surface observations were obtained every 12 h from 0000 on 14 August to 1200 on 20 August 1969. The radiosonde information was analyzed over the Gulf of Mexico, the Caribbean, and the United States at the standard pressure levels.

3. Results

a. *Rapid deepening phase*

The first significant event in the life of Camille was the rapid deepening which began before and continued after the cyclone center passed over the western tip of Cuba. An interpretation of the time history of the minimum central pressure curve (based upon U.S. Air Force reconnaissance aircraft reports) is shown in Fig. 3. There is a suggestion of two periods of deepening. The first primarily occurred between 1200 on 14 August and 1200 on 15 August. Four closely spaced reconnaissance aircraft reports centered at about 1200 GMT on the 15th all reported a minimum pressure of about 965 mb. Shortly thereafter the center crossed the southern coast of Cuba near the western tip. It took approximately 3 h for the eye to move to the northern coast. During this time, the shape of the minimum pressure curve is quite uncertain, but since most of the cyclone's circulation remained over the water it is likely that the rise, if any, in the central pressure was small.

[1] V. V. Salomonson, personal communication, 1972.

The next reconnaissance reports were made at 1800 GMT 16 August when Camille had reached severe status, an intensity that the aircraft central pressure measurements would indicate was maintained until the cyclone crossed the U. S. coastline. Thereafter, the central pressure rose rapidly. At Jackson, Miss., the minimum pressure was 978 mb as the center passed about 35 km east of the station at 1200 on the 18th.

Perhaps the best day and night satellite observations of Camille during the rapid deepening phase were made by the Nimbus 3 MRIR. By examining the information from the three IR channels at 12 h intervals and the visible channel data every 24 h, it was anticipated that the changes in the structure of Camille, as it became a severe cyclone, could be determined. The daytime observations were combined with the ATS 3 cloud motions to obtain a more complete picture of the circulation.

The first Nimbus 3 observation of Camille and the surrounding systems that will be examined was made between 0400 and 0800 GMT on 15 August. At this time, Camille was a tropical storm, rapidly approaching hurricane status about midway through the first deepening phase. Fig. 4 depicts the 10.5 μm measurements, when the storm had a prominent spiral band that extended to the east of the center and the highest clouds were near the center. There is also evidence of two other bands (best shown by the 260 K line) located well southeast of the center. The development of these bands is in apparent response to the organization of a broad southeast current from the surface to 700 mb which had a large inflow component. The surface data for 1200 on the 15th showed two reports of a southeast wind at 20 kt in the band area. At 850 mb, the wind speed increased to 30 kt. Thus, with the storm moving northwestward at 8 kt, this boundary-layer air was moving almost directly toward the storm center and was approaching the storm from that indicated area at an average speed of about 30 kt. This air was passing over water with a surface temperature of 303–304 K, which is 2 K above the mean sea surface temperature (about two standard deviations) for August in this region (U. S. Naval Oceanographic Office, 1967).

Northwest and east of Camille there are large regions where the 10.5 μm T_{BB}'s are $\geq$ 290 K (Fig. 4) which, when corrected for the water vapor effect of a tropical atmosphere by adding 6–8 K, are very close to the 302–304 K sea surface temperature reported by ships. The areas of warm 10.5 μm T_{BB}'s indicate that the amount and or the vertical development of low-level cumulus was being suppressed, probably due to subsidence in the lower troposphere. This method of interring subsidence has the limitation that once the clouds have been substantially suppressed or eliminated, there is no further information on the strength of the subsidence. However, because the 6.7 and 21.5 μm water vapor channels respond to the variation of water vapor in the upper and middle troposphere, respectively,

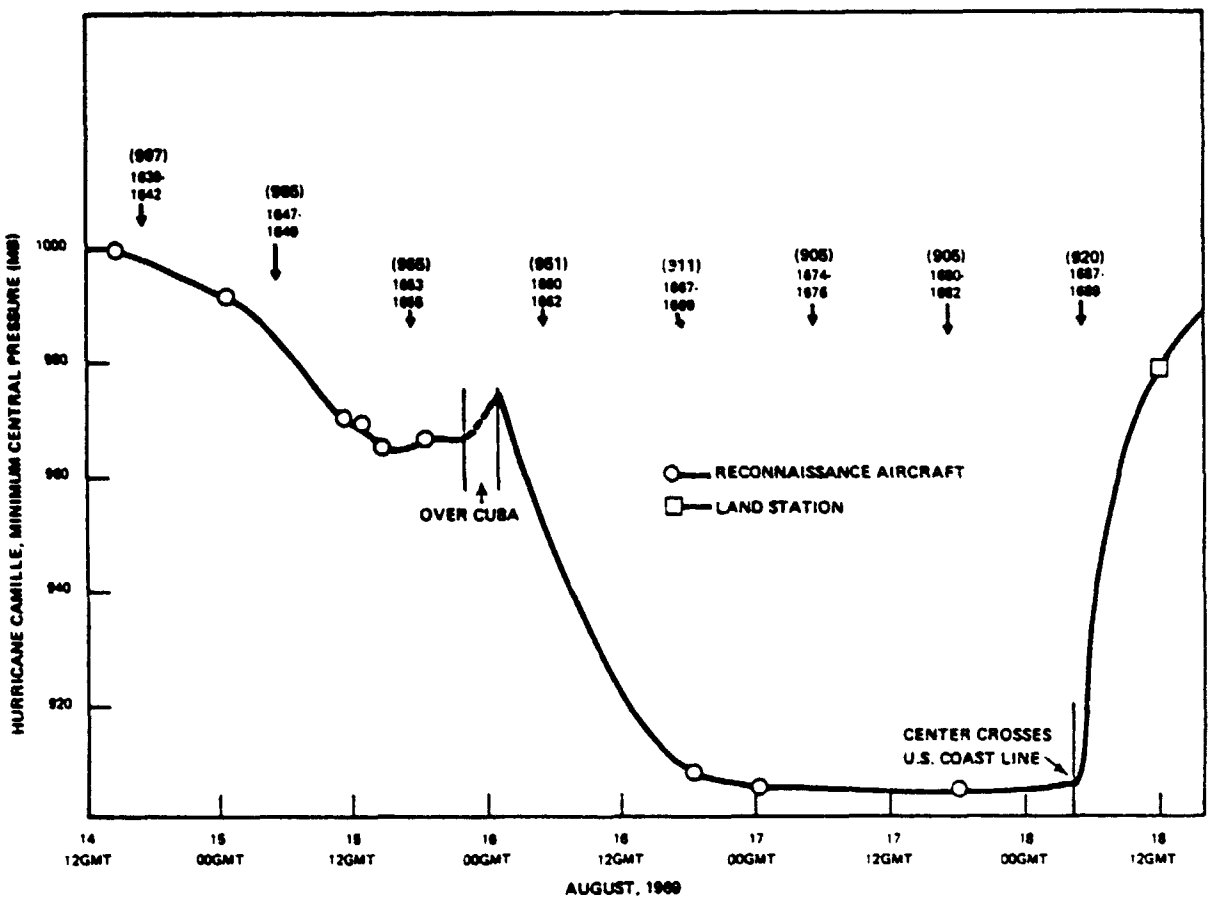


FIG. 3. A trace of the minimum central pressure for Hurricane Camille from 1200 GMT 14 August to 1800 GMT 18 August 1969. The estimated central pressures and the Nimbus 3 orbit numbers are shown above each of the arrows that indicate the time of a Nimbus 3 overpass.

subsidence may be inferred in the upper and or middle troposphere in areas of weak horizontal dry air advection when the corresponding T_{BB}'s increase (Rodgers *et al.*, 1976). The 6.7 μm T_{BB}'s (Fig. 5) within the area of the warm T_{BB}'s of the window channel were ≥ 240 K northwest of Camille and ≥ 235 K east of Camille. The area of warm 6.7 μm T_{BB}'s northwest of Camille was undergoing subsidence induced by horizontal speed convergence at 200 mb as revealed by a streamline and isotach analysis at that level. However, the horizontal dry air advection caused by the southerly displacement of this dry tongue from earlier observations made it more difficult to ascertain the strength of this subsidence. The warm 6.7 μm T_{BB}'s east of Camille were not high enough to indicate with reasonable certainty that cirrus clouds were not present. A comparison of high-altitude aircraft cloud photography and Nimbus 3 6.7 μm T_{BB}'s over the tropics has indicated that when the T_B range was 235–237 K, the cirrus probability was 40% (Shenk *et al.*, 1976).

The outflow north of the center was apparent from the widespread cirrus (where the 6.7 μm T_{BB}'s are ≤ 230 K) was widespread (Fig. 5). Camille was then in a moist environment, at least in the upper troposphere, as there was no evidence of dry air within 5° latitude from the center.

Twelve hours later there was evidence from the reconnaissance aircraft reports that the rate of deepening had slowed or stopped. The 10.5 μm channel chart with ATS-3 low-level cloud motions superimposed (Fig. 6) shows that Camille had a more circular appearance and that the area of intense cloudiness covered a larger area than 12 h earlier. There was an extensive area southeast of the center where the T_{BB}'s were ≤ 240 K, which was indicative of the continued cloud development in conjunction with the low-level confluence that was shown by the low cloud motions. The band of cloudiness was continuous from the ITCZ to Camille.

The development of the clouds in the confluence zone could be an important factor for a more rapid transport of moist air into the inner circulation of Camille. Conventional data indicate that the winds between the surface and 850 mb had about the same radial component toward the storm between 0000 on 15 August and 0000 16 August. The low-level winds 350 km south-

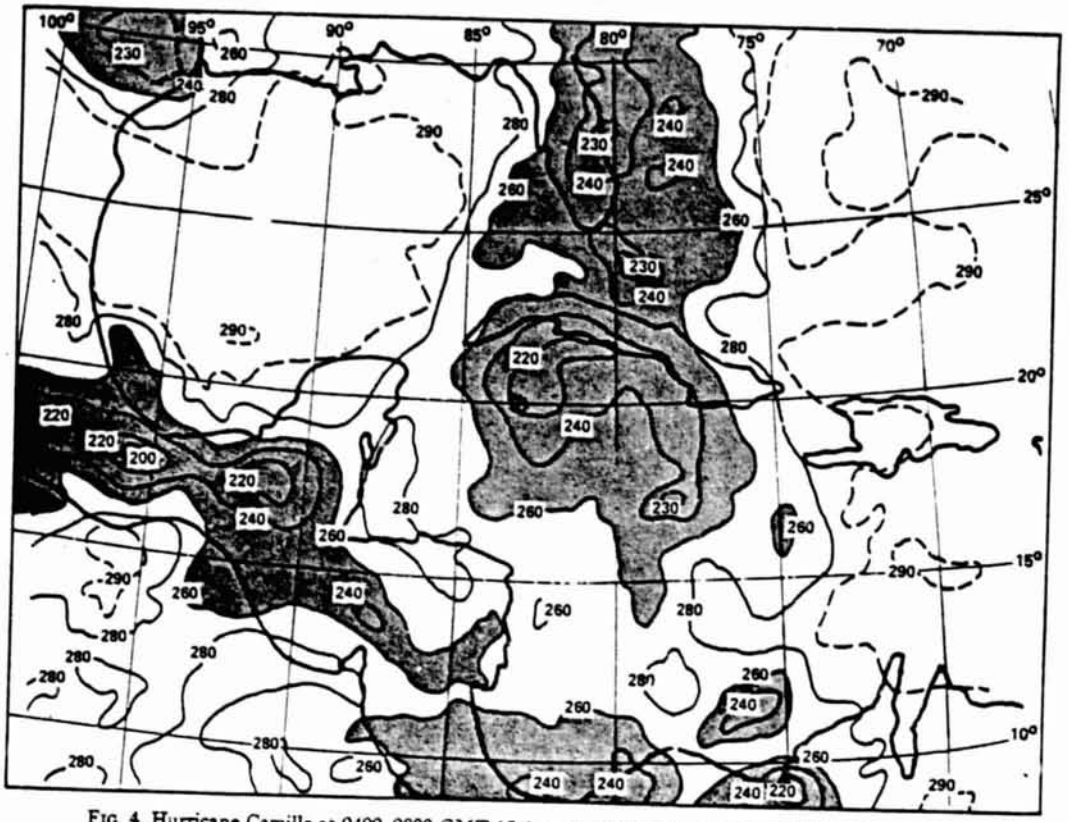


FIG. 4. Hurricane Camille at 0400–0800 GMT 15 August 1969, as depicted by 10.5 μm T_{BB}'s from the MRIR sensor. In these and all subsequent satellite images the isotherms are in kelvins and the cyclone center is indicated by the circled cross.

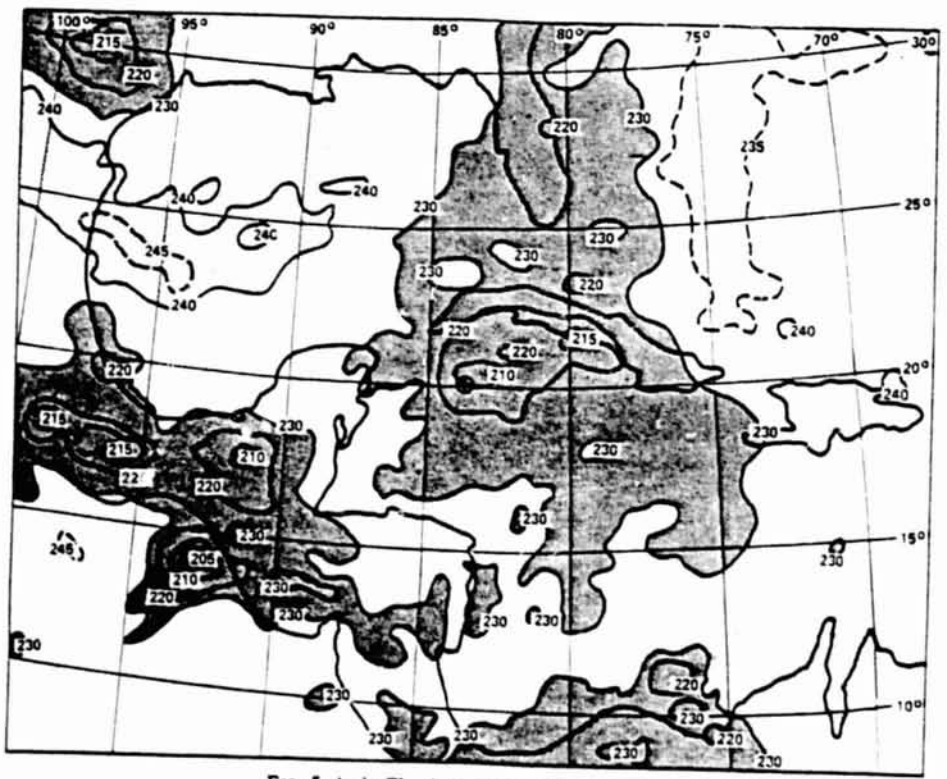


FIG. 5. As in Fig. 4 except for 6.7 μm T_{BB}'s.

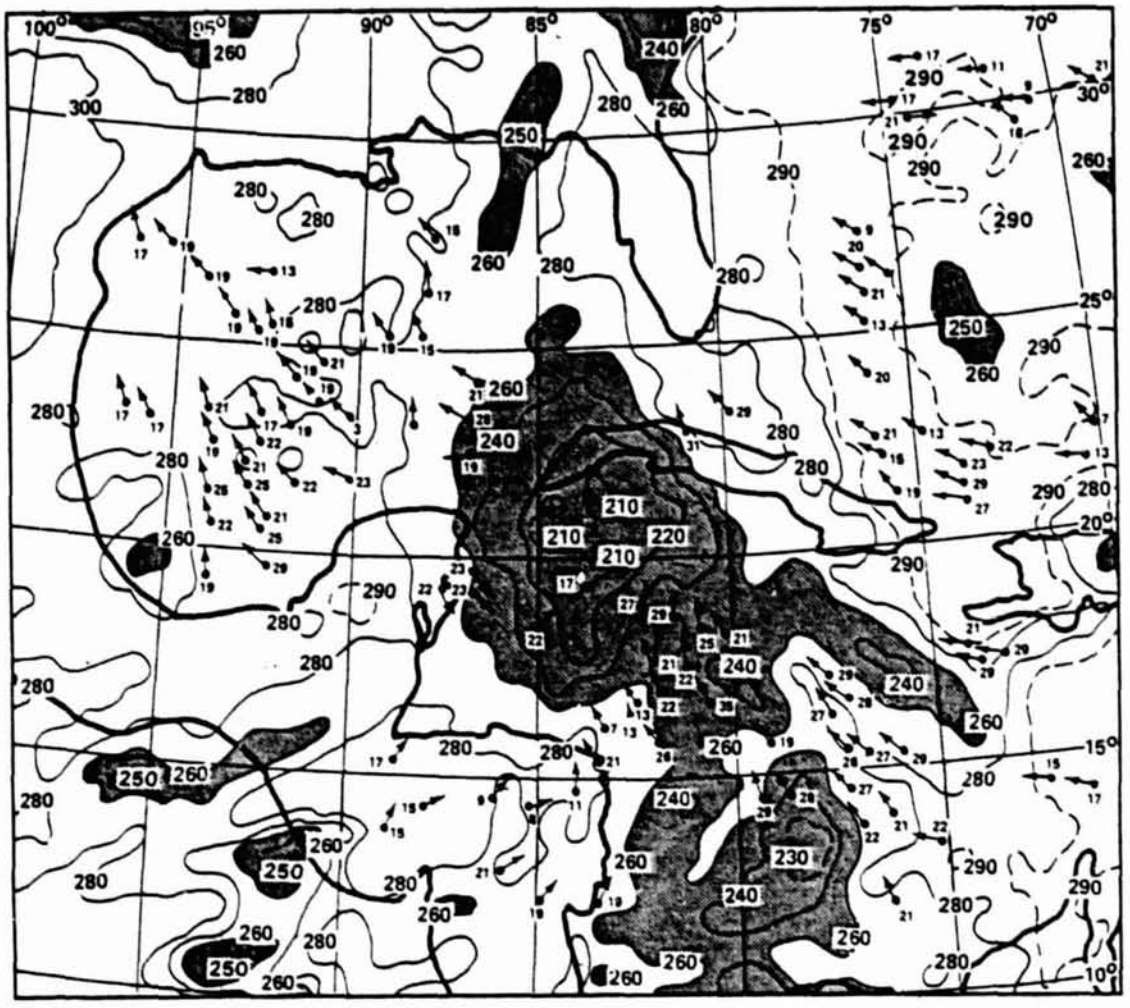


FIG. 6. Hurricane Camille at 1400–1800 GMT 16 August, 1959, as depicted by 10.5 μm T_{BB}'s from the MRIR sensor. Otherwise as in Fig. 4. Superimposed on the analysis of the radiation data is the surface analysis and cloud motion vectors derived from low clouds tracked from the ATS 3 satellite images. The cloud motion directions are indicated by the arrows and the speeds are in knots.

east of the center at 0000 on the 16th were as high as 36 kt, transporting moist air at 850 mb with a large radial component. These winds were preceded by 30 and 34 kt winds at similar locations 12 and 24 h earlier, respectively. Thus, with a favorable environment for 24 h, the cloud development was able to continue.

A small area of T_{BB}'s $\geq$ 290 K was located west of the storm, which indicated some suppression of low-level cloud development. To the east was a larger area that was most likely associated with upper horizontal convergence produced by the outflow, where the strength and position changed little on the 200 mb analysis between 1200 GMT 15 August and 0000 GMT 16 August. North of the cyclone center, the cirrus shield persisted, as seen in the 6.7 μm measurements (Fig. 7), indicating that ventilation is clearly shown by the middle and high cloud motions from ATS-3 where a strong southerly current exists at the cloud levels for at least 1100 km.

The 200 mb streamline and isotach analysis for 0000 GMT 16 August indicates that the flow over southern Florida had changed from southerly to westerly. This change was most likely an expansion of the outflow area of Camille. East of the cirrus there is a well-defined ridge line demarcated by 6.7 μm T_{BB}'s (as high as 244 K) and the cloud motions. The 200 mb analysis cannot confirm the presence of the line, since the high T_{BB}'s were east of Florida. It is likely that some of the cirrus ejected by the outflow dissipated in the dry air associated with the ridge line. West of Camille, the relatively high T_{BB} area had weakened and moved little with respect to the storm's center, and the upper horizontal convergence, as shown by the cloud motions, is coincident with the highest T_{BB}'s.

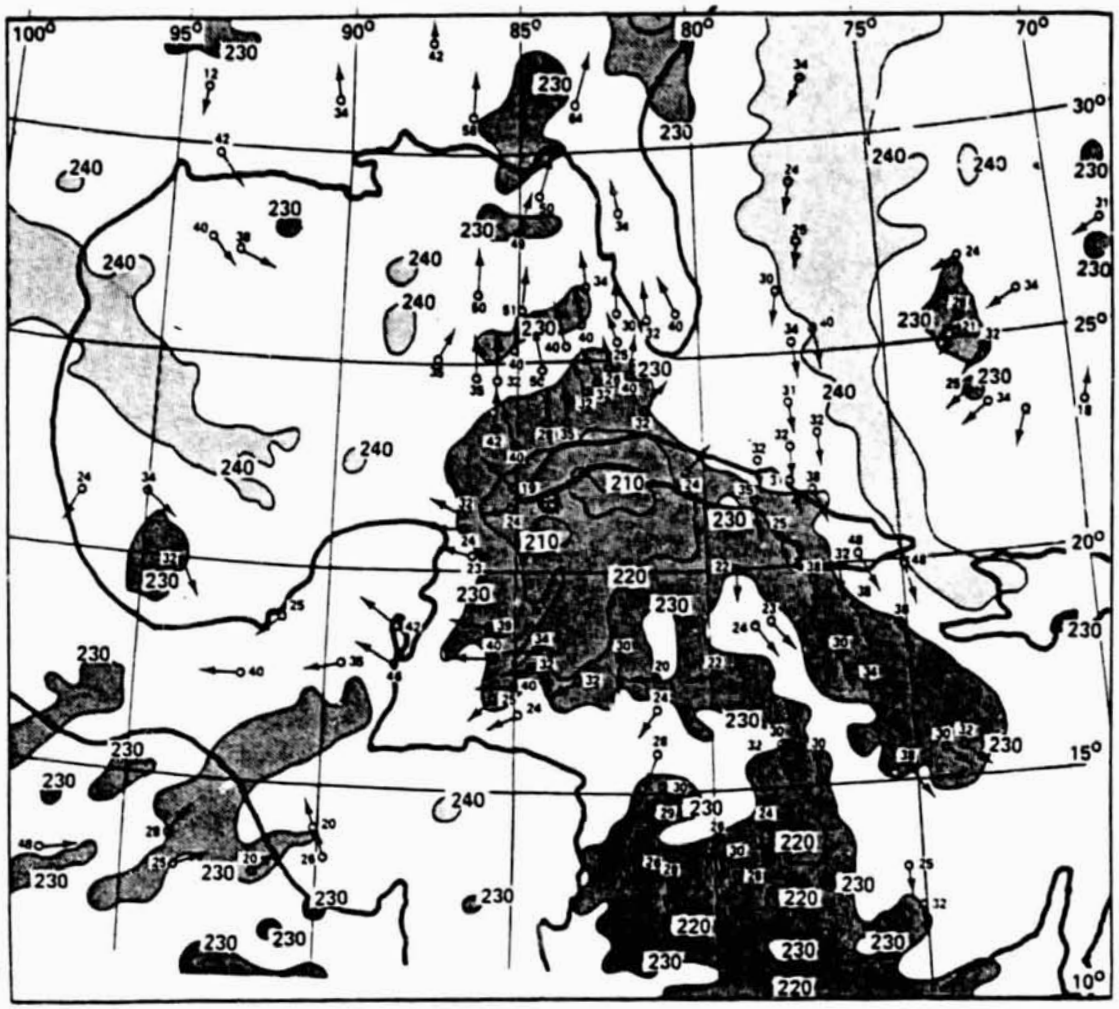


FIG. 7. As in Fig. 6 except for 6.7 μm T_{BB}'s. Superimposed on the analysis of the radiation data are cloud motion vectors derived from high and middle clouds tracked from ATS 3 satellite images and the streamline analysis based on the cloud motions and 200 mb rawins. The cloud motion directions are indicated by the arrows and the speeds are in knots.

Twelve hours later (0300–0700) on 16 August, Camille had probably begun the most dynamic deepening phase of its life cycle. The changes in the structure of Camille, as seen in satellite observations, also reached a climax. The 10.5 μm T_{BB}'s (Fig. 8) show that the size of the active cloudiness, as defined by the 240 K isotherm, had been reduced by about a factor of 3 from the last satellite observation. Therefore, during the early portion of rapid deepening, the cloud canopy contracted sharply. The cloud bands in the converging moist air southeast of the center were still strong and extended 1100 km from the center. The outer limit of the cloud bands was almost exactly the distance from Camille's center where the surface streamlines began to converge at 1200 on the 16th. There were no rawinsonde measurements close to the center, but the surface reports show the continued inflow where the wind speeds were 20–30 kt. West of the center, the area covered by $T_{BB} \geqslant 290$ K had expanded considerably in 12 h. This large expansion represents a substantial increase in the size of the region where the cloudiness was being suppressed. This would indicate that the subsidence surrounding Camille over a large area was beginning to affect the cloudiness in response to the increased vigor of the inner circulation.

The 6.7 and 21.5 μm measurements (Figs. 9 and 10) also indicate the subsidence increase west and northwest of the center at higher levels. Twelve hours earlier there had been a weakening band of high T_{BB}'s oriented northwest–southeast that was most likely advected from north of the storm. The high T_{BB}'s at this time encircled the center in the western semicircle, which suggests that they were produced by the hurricane. They cover approximately the same area as that covered by the $\geqslant 290$ K isotherm in Fig. 8. It is interesting that the time lag between the increased vertical motion in the inner portion of Camille and the subsidence appeared to be small. The cyclone center had

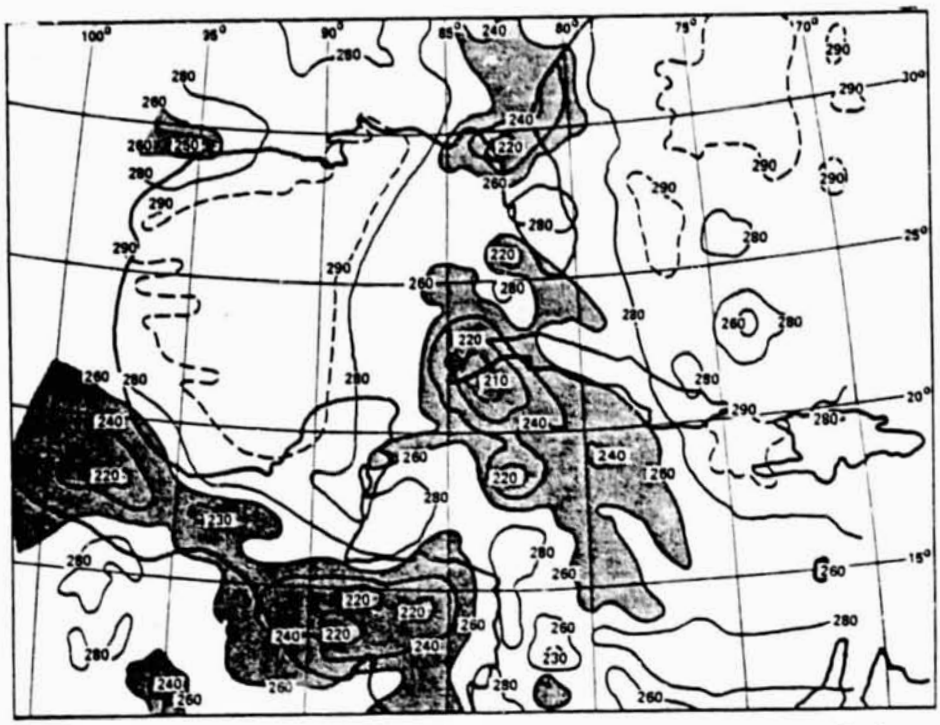


FIG. 8. Hurricane Camille at 0300–0700 GMT 16 August 1969, as depicted by 10.5 μm T_{BB}'s from the MRIR sensor.

moved off the northern Cuban coast for only about 4 h. A more exact assessment of the lag would require geosynchronous satellite measurements.

Fig. 11 depicts the cloud shield of Camille at 0500 GMT 16 August with HRIR measurements (1:2 000 000 map scale) 4 h after the center moved off the Cuban

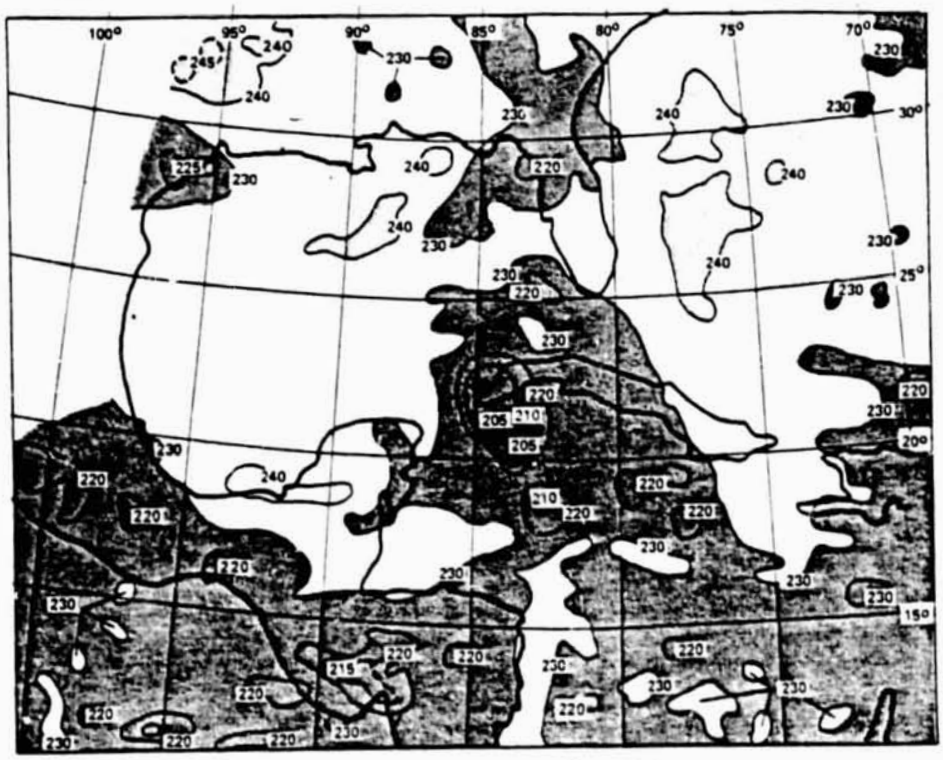


FIG. 9. As in Fig. 8 except for 6.7 μm T_{BB}'s.

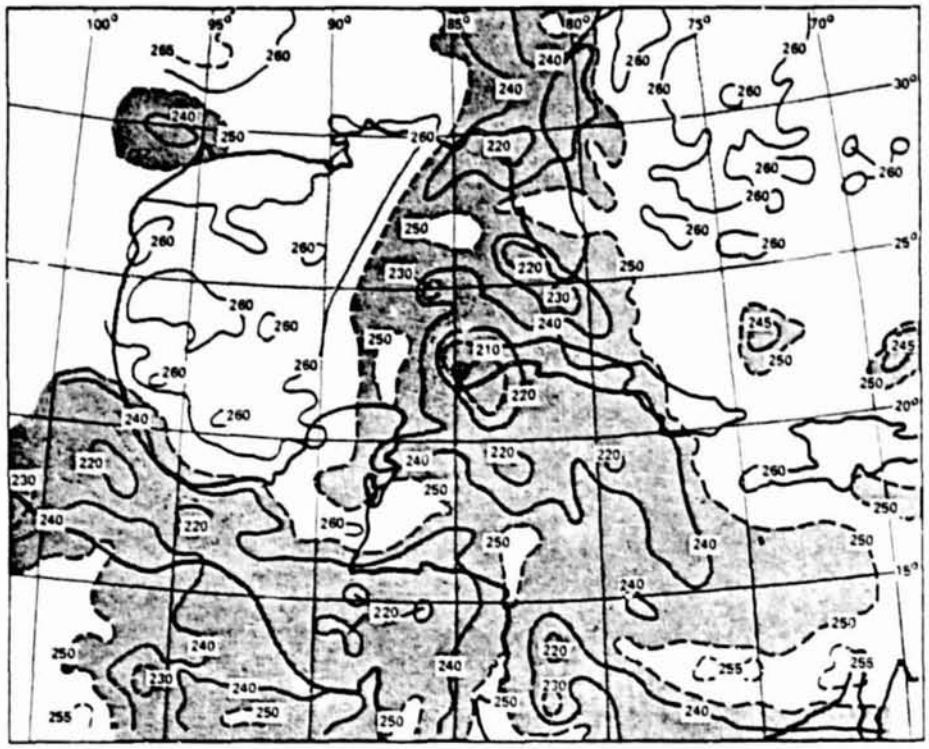


FIG. 10. As in Fig. 8 except for 21.5 μm T_{BB}'s.

Coast. Two features suggest that further deepening had begun. First, the appearance of a well-defined eye is indicated by T_{BB}'s as high as 287 K (the insert in Fig. 11 depicts a small area near the center mapped at 1:500 000). This measurement could have been caused by either nearly transparent cirrus and or low clouds, or a partially filled radiometer field of view with any of the above cloud combinations. Second, a ring of very low T_{BB}'s ($\lesssim$ 210 K), indicating the opaque storm cloudiness, surrounds the eye except to the west. This ring represents the highest clouds associated with the wall cloud where the lowest (and probably the weakest) portion is on the west side.

By 1500–1900 on the 16th, Camille had become a very severe hurricane. At 1800, a U. S. Air Force reconnaissance aircraft measured a central pressure of 908 mb. Therefore, Camille deepened about 60 mb in 24 h. The 10.5 μm measurements (Fig. 12) show that the intense portion of the Camille cloud canopy ($T_{BB} \lesssim$ 240 K) had expanded by about a factor of 4 in 12 h. This expansion (and the contraction 12 h earlier) are not attributed to a diurnal oscillation but are hypothesized to be manifestations of Camille's intensity change. A minimum (maximum) tropical cyclone cirrus cover has been observed from the SMS infrared imagery analysis of 16 days of eight tropical cyclones to occur at approximately 0300 LST (1700 LST) (Browner *et al.*, 1977) 7–8 h between the Nimbus 3 observation times. There were two areas of great activity, one near the center with T_{BB}'s between 200 and 204 K (the lowest recorded during the entire life cycle) and another of 205–209 K imbedded in the broad band southeast of the center. The low T_{BB} measurements near the eye were almost perfectly centered, thus suggesting an intense circular wall cloud. This is verified by the concurrent high-resolution (3 km) Image Dissector Camera System (IDCS) image (Fig. 13) in which the eye is clearly evident. The wall cloud is seen as a bright inner ring surrounded by a cloud system of lower brightnesses than the clouds which formed the rest of the canopy or the wall cloud. These features suggest that the cloud tops within the wall cloud ring had reached a higher level than the surrounding clouds, and that the air was descending immediately once it moved outside of the wall cloud. The subsidence would dissipate the cirrus or reduce its opacity and produce the lower brightness ring that separates the wall cloud from the rest of the cloud system.

Southeast of the center, the broad cloud band system was continuous from the ITCZ into Camille's circulation, and the low 10.5 μm T_{BB}'s seen in Fig. 12 indicate that a substantial percentage of the area contained middle and high cloudiness. The low-level cloud motions shown in Fig. 12 continue to exhibit the general confluence in the low-level wind field, with the cloud speeds toward the cyclone much higher than the northwest motion of the storm. West of the center, the area of T_{BB}'s $\gtrsim$ 290 K had continued to expand and measure-

ments of 295 K covered approximately 40% of the area where T_{BB}'s of ≥ 290 K occurred over the Gulf of Mexico. Thus the suppression of cloudiness (probably low level) had become more widespread and had intensified in some of the areas to the point where all cloudiness had probably disappeared. After correction for the atmosphere, a 295 K T_{BB} was within 2 K of the sea surface temperature. The cloud motions, seen in Fig. 12, showed evidence of low-level divergence over portions of the area where $T_{BB} \geq 290$ K. Also, the 0.2–4.0 μm reflectance (Fig. 14) of $\leq 10\%$ covered nearly the entire Gulf west of the Yucatan Peninsula.

The water vapor channels exhibited some remarkable changes (Fig. 15 and 16). In the entire western semicircle over a radius of about 5° latitude, higher T_{BB}'s were found with the maximum measurements of 250 K in the 6.7 μm channel northwest of the center and 270 K in the 21.5 μm channel further away and west of the eye. Massive subsidence can be inferred as the increased circulation within Camille became more evident as reflected in the large ejection of air to the periphery. This is particularly impressive when the evidence from the conventional data indicates that the 200 mb flow east of the storm also increased between 0000 and 1200 on 16 August and did not diminish in the following 12 h. The region of maximum upper tropospheric convergence appears to be along a line 5° latitude northwest of the center, as seen in Fig. 16, where the 6.7 μm T_{BB}'s were ≥ 240 K. Air was apparently flowing northwestward to converge with the eastward moving air around the base of the trough over the central United States. A combination of the 200 mb reports for 1200 GMT and the high and middle cloud motions from ATS supports that interpretation of the radiation map (Fig. 16). The analyzed line of upper horizontal convergence was within 100 km of the area of ≥ 240 K T_{BB}'s northwest of the cyclone center. Because most of the emission in the 6.7 μm channel came from the 300–500 mb layer, which was just below the region of probable maximum horizontal convergences at about 200 mb, the subsiding air in this layer was most likely not far from the position where the air initially began its descent. The location and shape of the maximum 21.5 μm T_{BB}'s, as seen in Fig. 15, suggests that the air did not descend vertically. The maximum descent in the 400–700 mb layer is inferred to be further west and over a larger region. With the lack of conventional measurements, it was not possible to correlate the 12 or 24 h 6.7 μm and 21.5 μm T_{BB} changes with computed vertical motions. However, these T_{BB} changes should be related to the relative difference between the vertical motion in the two layers. Over the area 6–8° latitude west of the center, where clouds were probably absent for the 24 h interval, the ΔT_{BB}'s in the 6.7 and 21.5 μm channels were about 6 and 10 K, respectively. Thus, a rough estimate of the ratio of the subsidence strength would be 10/6, with the greatest sinking in the 400–700 mb layer. Since the maximum vertical motion usually occurs at the level of nondivergence, which is about 600 mb, this result was not surprising.

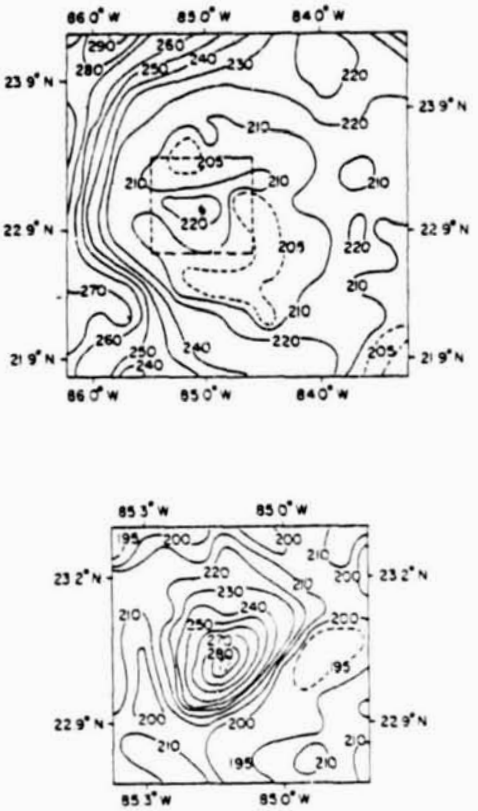


FIG. 11. Two views of the interior portion of the cloud structure of Hurricane Camille at 0500 GMT 16 August 1969, as depicted by the 3.8 μm equivalent blackbody temperatures (T_{BB}'s) from the Nimbus 3 HRIR. The top picture was mapped at a 1:2 000 000 scale, while the insert over the "eye" seen in the bottom picture (the area within the dashed box in the top picture) was mapped at a 1:500 000 scale.

As Camille continued to move toward the Mississippi coast, it appeared somewhat weaker in all the MRIR measurements although according to the aircraft monitoring Camille it did not diminish in intensity until it crossed the coast. The major MRIR observed changes were a reduction in size of the cirrus canopy and the cloud band southeast of the center. In addition, the subsidence west of the center appeared to be weaker. A probable major contributing factor to Camille maintaining its intensity was that the ship recorded sea surface temperatures surrounding the storm were at 2–3 standard deviations above normal (304–.05 K). Also, it is possible that the weaker appearance in the satellite measurements represented a recovery from the earlier surge that went past a substainable level.

b. Interaction with the westerlies

Camille rapidly weakened as it moved inland. By 1200 on the 18th the storm had reached Jackson, Miss.,

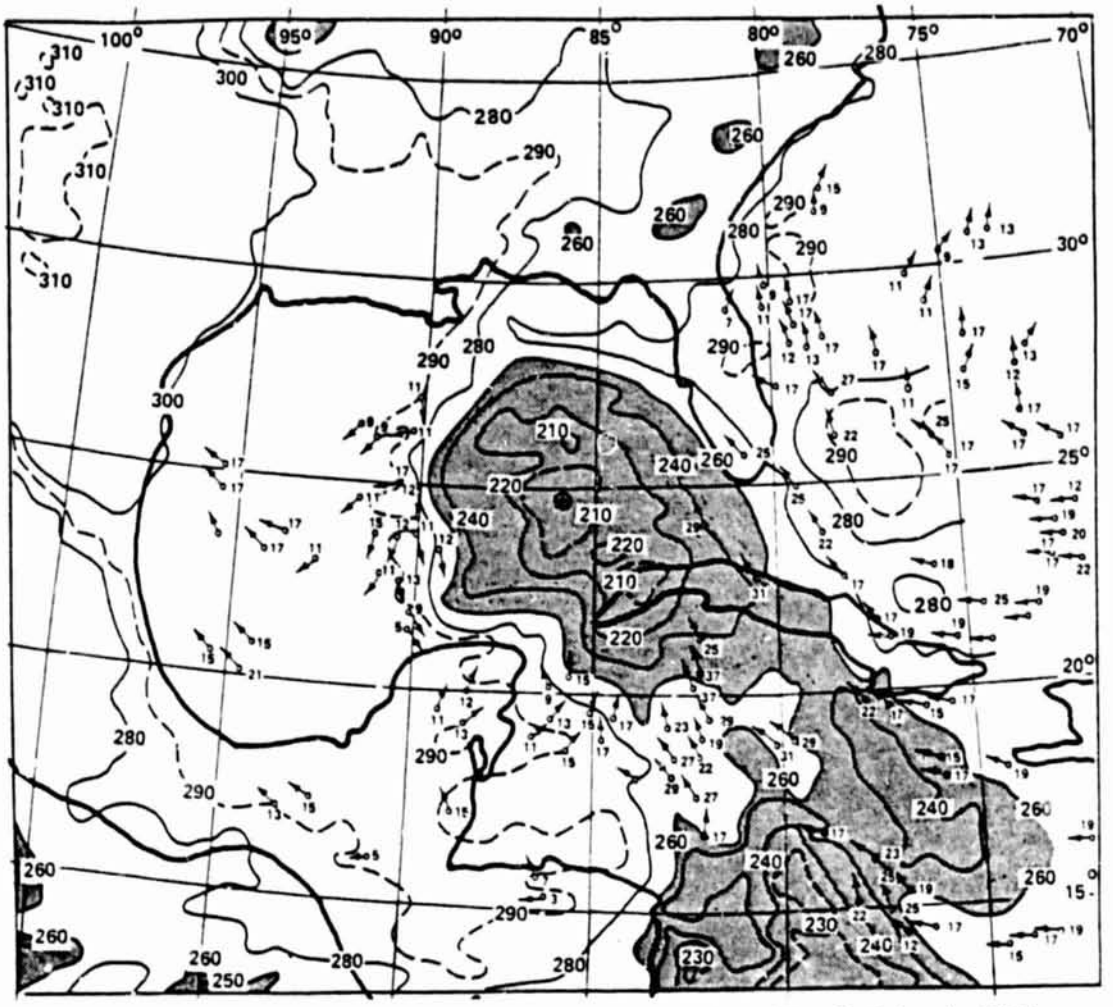


FIG. 12. Hurricane Camille at 1500–1900 GMT 16 August, 1969, as depicted by 10.5 μm T_{BB}'s from the MRIR sensor. Superimposed on the analysis of the radiation data are the surface analysis and cloud motion vectors derived from low clouds tracked from the ATS 3 satellite images. The cloud motion directions are indicated by the arrows and the speeds are in knots.

with maximum wind gusts of about minimum hurricane force. Nimbus 3 observed the cyclone near midday when Camille was at tropical storm strength. There had been a sharp decrease in the maximum cloud-top levels as evident from the 10.5 μm T_{BB} seen in Fig. 17. The minimum 10.5 μm T_{BB} is 225 K, and assuming that the clouds were opaque and filled the field of view of the radiometer, this temperature coincides with the 200 mb level of the 1200 sounding for Jackson, Miss. Twelve hours earlier, the minimum cloud top T_{BB}'s were 205–209 K, which is near 150 mb on the 0000 GMT 18 August soundings, both at Jackson and Shreveport, La. The 2 km drop in maximum cloud-top height was most likely due to the diminished vigor of the upward vertical motion in the most intense activity near the center.

Perhaps the most interesting satellite observations made of Camille during this phase of its life cycle were taken by the two water vapor channels (Figs. 18a and 18b). Surrounding the center to the north and west was a band of high T_{BB}'s which were higher than at 12 h earlier. This was surprising, since the cyclone had weakened so much, and therefore, it should follow that any storm-produced subsidence would diminish.

To examine this area of warm water vapor T_{BB}'s which was observed at 1200 on the 18th and again at 0000 on the 19th, a quasi-geostrophic adiabatic 10-level diagnostic model (Barr *et al.*, 1970) was employed in order to investigate the middle and upper tropospheric dynamics. There are two areas that were examined with the model. The first was located behind a weak upper tropospheric trough just north of Camille, and the second was an area of horizontal convergence between Camille's outflow and the upper tropospheric environmental flow northwest of Camille. Both of these areas appeared in the tongue of warm T_{BB}'s measured in the water vapor channels north and northwest of Camille at 1800 GMT on 18 August. Four three-dimensional trajectories were computed at 3 h intervals between

1200 on the 18th and 0000 on the 19th) at the 350 and 550 mb levels, in order to examine the warm 6.7 and 21.5 μm T_{BB}'s, respectively (Figs. 18a and 18b). The trajectories at both levels originated at the same location.

In Fig. 18a, trajectory 3 shows an air parcel moving through the base of the trough. During the first 9 h, the parcel experienced subsidence induced mainly by differential advection of negative vorticity from behind the trough. Equivalent blackbody temperatures of less than 245 K correlate with this area of sinking air. However, after subsiding 16 mb, the air parcel moved into an area of ascending motion in front of the trough.

In the middle troposphere, the parcel of air, as represented by trajectory 3, was less influenced by the trough and more influenced by an area of divergence north of Camille. During the first 3 h, the air parcel subsided as it moved northeastward but then ascended during the remaining 9 h as the air parcel moved ahead of the trough. Cross section AA in Fig. 19a, which is located in a west–east direction vertically through the base of the trough (designated by line AA in Figs. 18a and 18b), delineates the vertical extent of the subsidence which affected the air parcel flowing through the base of the trough represented by trajectory 3. It can be seen that the greatest subsidence occurred in the upper troposphere at 92° W. This subsidence increased the temperature-dewpoint spread in the troposphere down to 900 mb, as observed in the vertical distribution of specific humidity.

FIG. 13. Hurricane Camille at 1900 GMT 16 August 1969, as recorded by the Nimbus 3 Image Dissector Camera System.

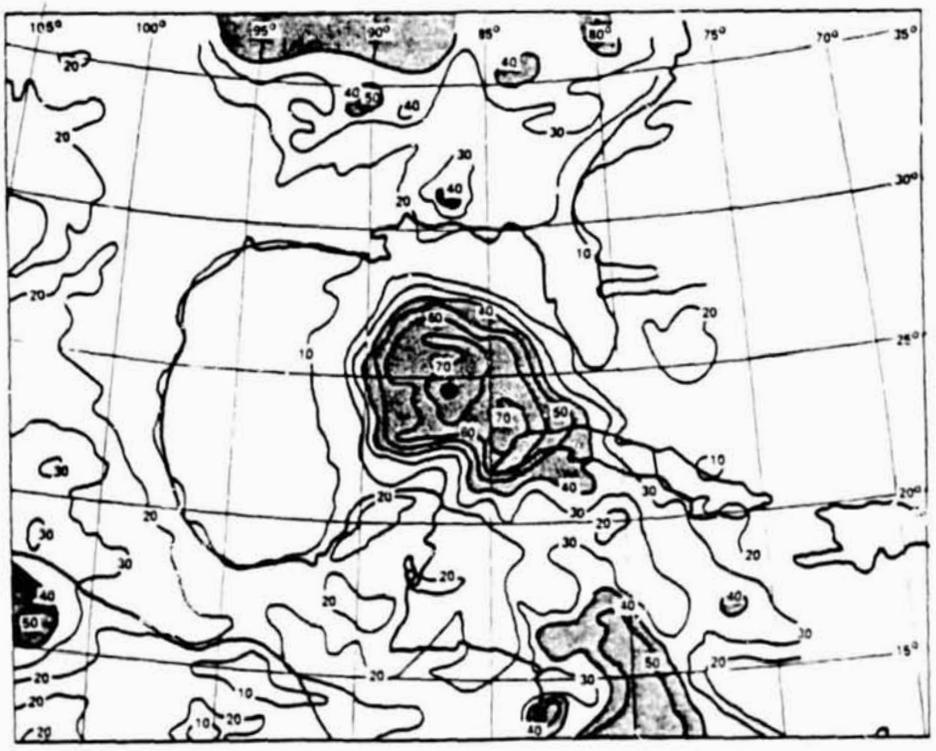


FIG. 14. Hurricane Camille at 1500–1900 GMT on 16 August 1969, as depicted by 0.2–4.0 μm normalized reflectance measurements from the MRIR sensor (spectral albedo in percent). The cyclone center is indicated by the circled cross.

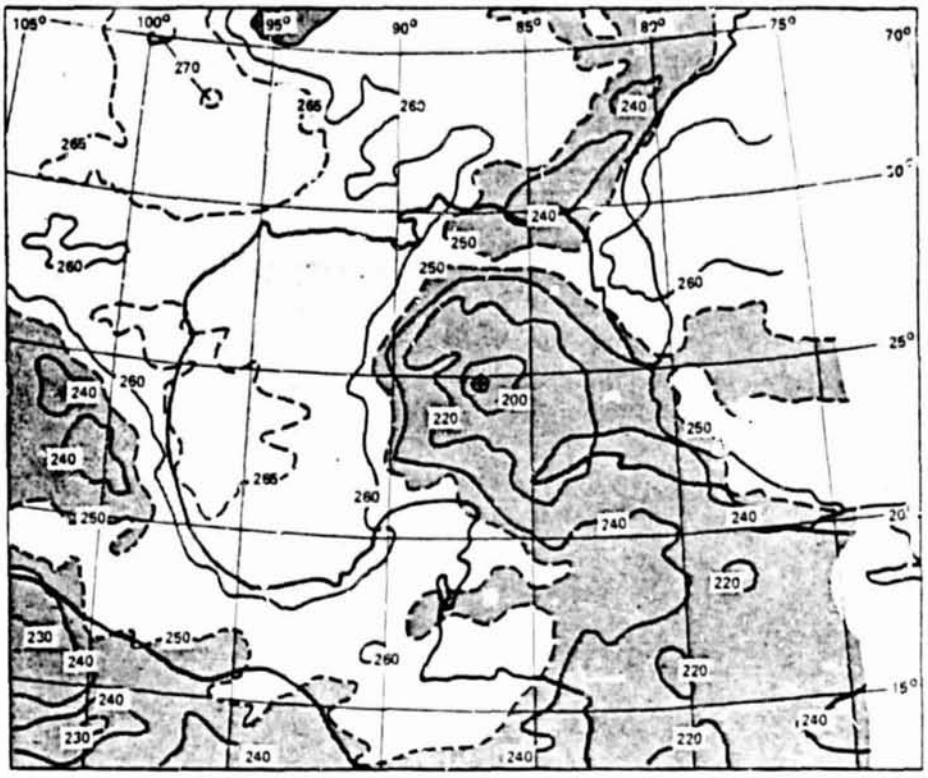


FIG. 15. As in Fig. 14 except for 21.5 μm T_{BB}'s from the MRIR sensor.

In the second area, it was found from vertical mass influx computations at all 10 levels for 1200 GMT 18 August that the maximum horizontal convergence of $\leqslant -5 \times 10^{-4}$ s^{-1} was above the 350 mb level with horizontal divergence below. Trajectories 1, 2 and 4 in Figs. 18a and 18b delineate the flow pattern in this area of the middle and upper troposphere. At the 350 mb level, trajectories 1 and 2 (Fig. 18a) best outline the upper tropospheric flow from the base of the trough southwestward into the area of convergence. It is seen that the parcels were first under the influence of subsidence behind the trough, but as they moved out of the trough and into the convergence area, the parcels continued to subside due to the enhanced subsidence caused by the upper level convergence. Further downstream, southwest of the convergence area outlined by trajectory 4, the air parcel continued to subside during the first 6 h as it moved southwestward, but at a much slower rate. This subsidence was mainly induced by the horizontal advection of colder air from the northeast. After 6 h, the parcel moved into an area of ascending motion. In this area of descending air, the 6.7 μm T_{BB}'s were $\geqslant$ 245 K. At the 550 mb level (Fig. 18b), the parcel in trajectory 1 descended nearly 12 mb in the first 6 h, and 8 mb during the last 6 h, indicating that air moved out of the trough at this level had continued to subside under the influence of forced subsidence caused by the horizontal convergence aloft. Trajectories 2 and 3 (mentioned previously) illustrate the middle tropospheric divergence north of Camille. The diverging air subsided as depicted by the two trajectories. Trajectory 4 again outlines the subsidence further downstream southwest of Camille. Here, under the influence of horizontal cold-air advection, the air continued to subside throughout the period. The 21.5 μm T_{BB}'s in this area of subsidence were $\geqslant$ 255 K. Cross-section BB in Fig. 19b, which is located west-east through the convergence area (designated by line BB in Figs. 18a and 18b), again delineates the vertical extent of the subsidence. Between 88° and 98° W, the subsidence extended throughout the troposphere with weaker subsidence in the upper troposphere and stronger subsidence in the middle troposphere. The vertical axis of the subsidence tilted eastward with height. Again the vertical column was drier within the subsidence area as depicted by the specific humidity lines. Thus, the trajectories and vertical cross section suggest that there were two tongues of subsidence, one moving cyclonically northeastward through the base of the trough and the other moving anticyclonically southwestward out of the trough and into the area of enhanced subsidence below the convergence area. Therefore, since the 6.7 and the 21.5 μm water vapor channels respond to the variation of water vapor in the upper and middle troposphere, respectively, the increased drying of the atmosphere caused by the enhanced subsidence would

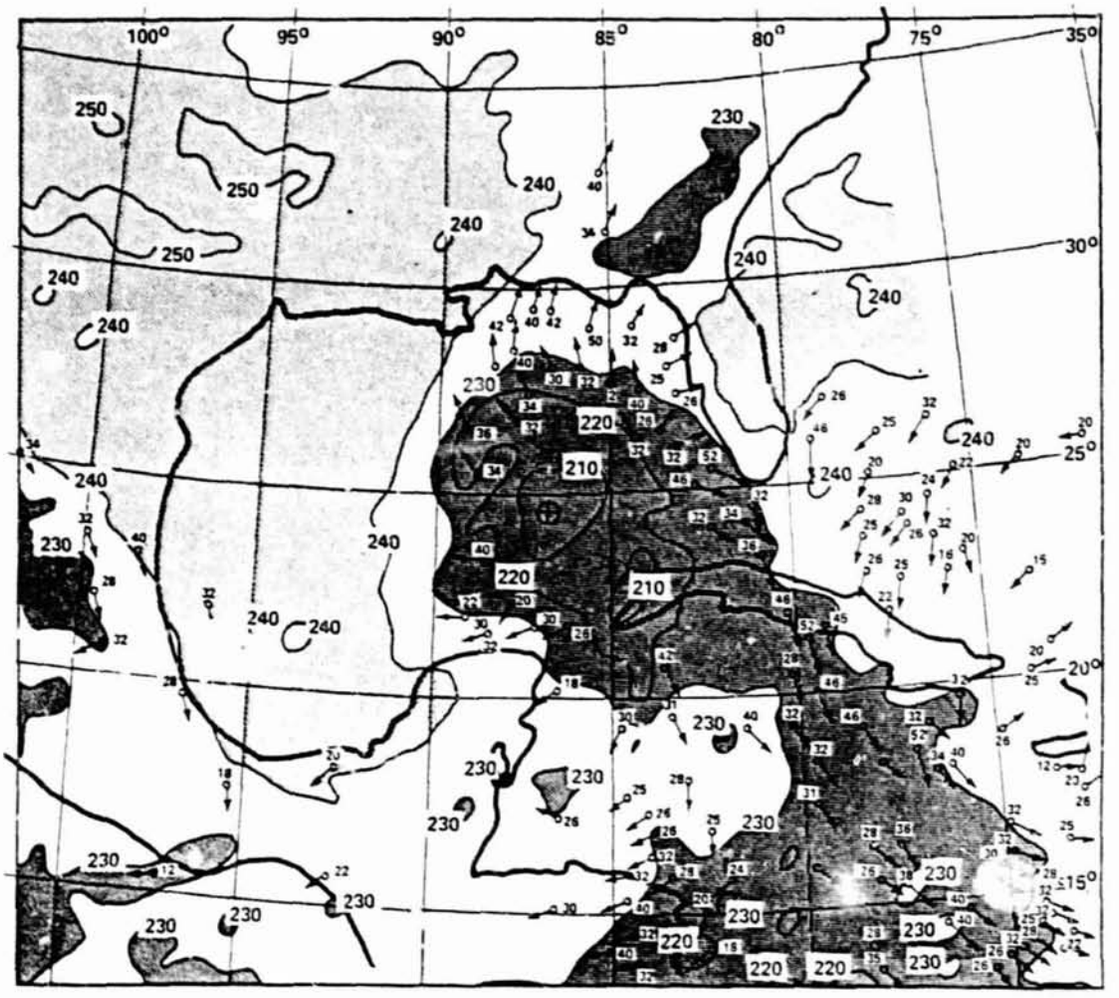


FIG. 16. As in Fig. 15 except for 6.7 μm T_{BB}'s. Superimposed on the analysis of the radiation data are cloud motion vectors derived from high and middle clouds tracked from the ATS 3 satellite images and the streamline analysis based on the cloud motions and 200 mb rawins. The cloud motion directions and speeds are indicated by the arrows and the speeds are in knots.

be reflected by the warmer water vapor T_{BB} measurements.

The radiosonde measurement at Nashville, taken at 0000 GMT 19 August showed that Camille had weakened further by the time the cyclone had reached extreme northern Mississippi. By then, Camille had been downgraded to a tropical depression. The strong convergence area in the upper troposphere seen 12 h earlier had almost disappeared, due to an anticyclone west of the center building eastward with a small ridge north of the storm center. The dynamics of the upper and middle troposphere had changed. With Camille weakening further and becoming more embedded into a more zonal westerly flow, the interaction between Camille and the westerlies had become weaker. Calculations at all 10 levels in the model revealed that there was still an area of upper tropospheric horizontal convergence north of Camille, but with a weaker magnitude. Cross-section AA (Fig. 20a)[3] indicates that the subsidence between 90° and 97° W that was attributed to differential advection of negative vorticity from behind the trough was weaker than that 12 h earlier. The vertical extent, however, still extended throughout the troposphere which continued to keep the column dry as seen in the isolines of specific humidity. Cross-section BB (Fig. 20b)[3] reveals that the subsidence between 90° and 95° W, associated with the upper tropospheric horizontal convergence between the westerlies and Camille's outflow, had weakened. However, the subsidence continued to keep the column dry as shown by the specific humidity isolines. The strong subsidence further west at 98° was not related to the Camille interaction with the westerlies, but was induced

[3] Cross sections AA and BB in Figs. 20a and 20b intersect the same areas as in Fig. 18.

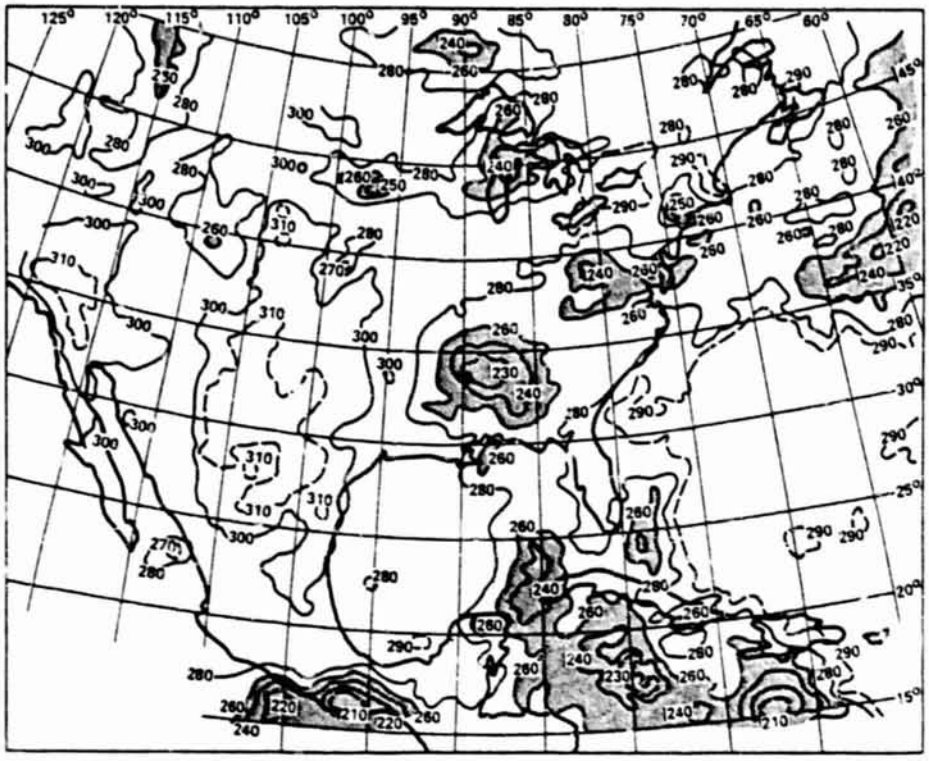


FIG. 17. Hurricane Camille at 1600–1900 GMT 18 August 1969, as depicted by 10.5 μm T_{BB}'s from the MRIR sensor.

by horizontal cold air advection from around the ridge west of Camille.

c. *Excessive rains over the central Appalachians*

As Camille moved northeast and then east across western Tennessee and Kentucky, it was a tropical depression producing moderate to locally heavy rainfall of 5–10 cm at stations close to the track of the center. However, as the storm center reached West Virginia, heavy rainfall began in west central Virginia around 0000 GMT on the 20th and continued for approximately 12 h. Schwarz (1971) has reported that the extreme rainfalls were associated with near record low-level moisture that was not subject to depletion by upwind mountain ridges. The Appalachians contributed to the large local amounts of rainfall (as high as 69 cm) because the low-level flow was from the southeast, perpendicular to the mountain range. Upper tropospheric horizontal divergence is another important condition that would favor regeneration of the cloud mass and lead to the production of heavier precipitation. A strong field of horizontal speed and direction divergence was present just northeast of the cyclone center, at 200 mb.

The regeneration of the cloud mass was best seen by the Nimbus 3 HRIR which observed Camille's cloud canopy at 24 h intervals for two days before and during the excessive rain period.

Fig. 21a provides a standard of reference for the two later HRIR views of Camille as the cyclone was moving ashore on the Mississippi coast. The lowest 3.5–4.1 μm T_{BB}'s are $\leqslant 210$ K over a small area near the center. A maximum cloud-top height of 125 mb (15.4 km) can be inferred from the minimum T_{BB} and the 0000 Lake Charles temperature profile on 18 August.

Twenty-four hours later, the greatly weakened cyclone was over western Tennessee. The HRIR view (Fig. 21b) indicates that the cloud pattern was more disorganized with the highest clouds (lowest T_{BB}'s) about 300 km northeast of the center. There had been an increase in the minimum T_{BB} to about 215 K. Using the 0000 GMT 19 August sounding at Nashville, this T_{BB} measurement would mean the maximum cloud top was at 180 mb (13.2 km). Thus, there was an apparent reduction in the maximum cloud top height of 2.2 km in 24 h.

The third and final HRIR view of the series (Fig. 21c) is at the time of the excessive rainfall. Most of the intense cloudiness (as outlined by the area where the T_{BB}'s were $\leqslant 230$ K) was north and east of the center similar to the position of the cloud shield for an extratropical cyclone. There were several areas where the T_{BB} values were $\leqslant 210$ K with a minimum of 203 K.

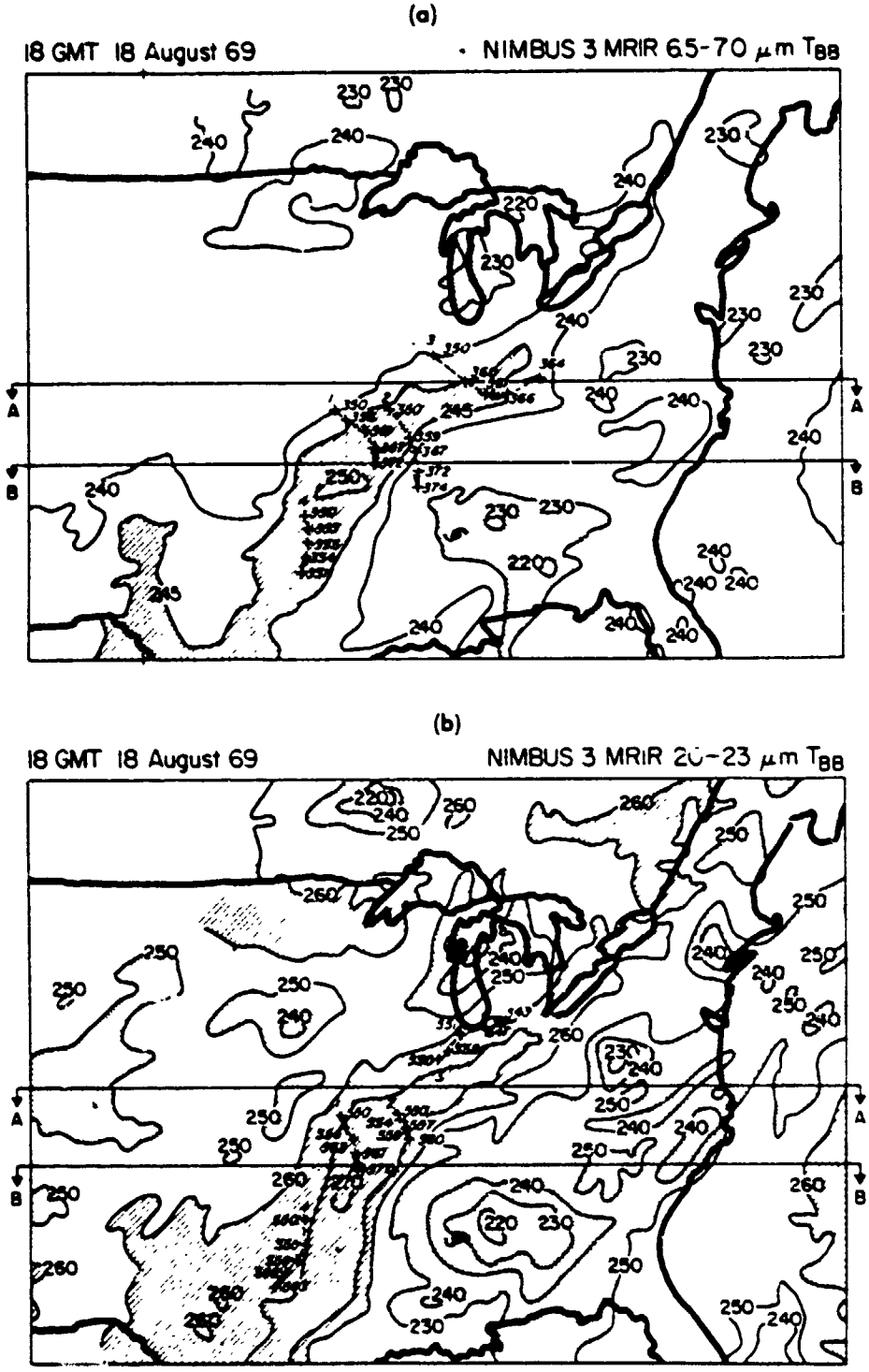


FIG. 18. Three-dimensional trajectories between 1200 GMT 18 August 1969 and 0000 GMT 19 August 1969, initiated at (a) the 350 mb level superimposed upon the 6.7 μm T_{BB}'s and (b) the 550 mb level superimposed upon the 21.5 μm T_{BB}'s at 1600–1900 GMT 18 August 1969 from the MRIR sensor. The crosses represent the position of the air parcels at 3 h intervals and the number to the side represents the pressure height. Lines AA and BB denote cross sections AA and BB.

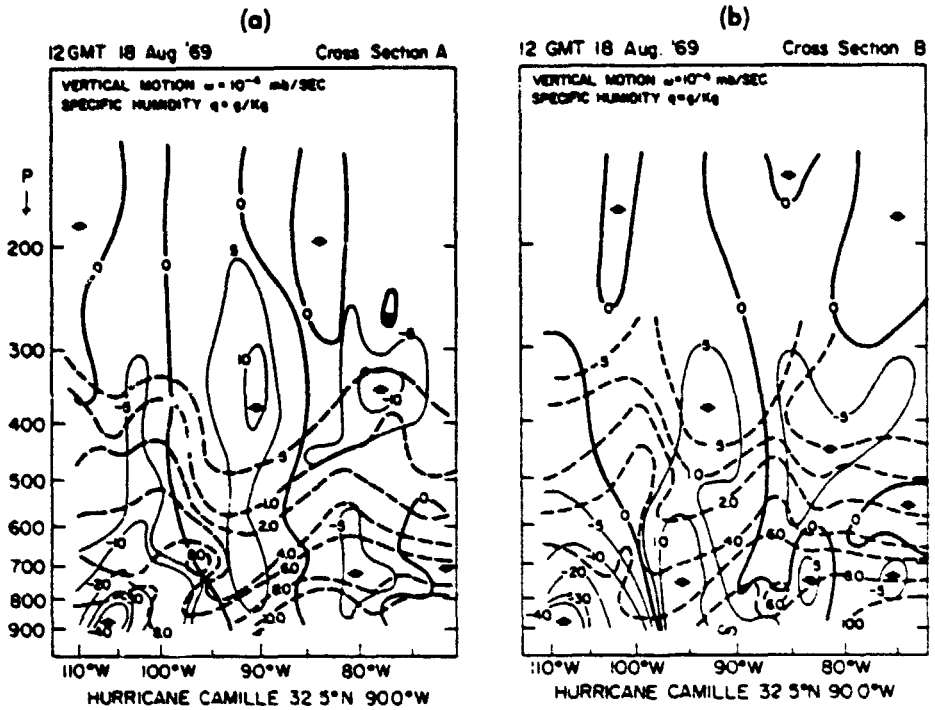


FIG. 19. Cross sections along (a) AA and (b) BB of vertical motion (solid lines) and specific humidity (dashed lines) for 1200 GMT 18 August 1969.

Therefore, the maximum cloud top height had reached the 125 mb (15.4 km) level which was the maximum cloud-top level when Camille was still a severe hurricane, shortly after crossing the Mississippi coast.

Extratropical secondary storm development has occurred when there was a significant vertical cloud growth within a large existing cloud mass; this growth can precede the detection of the new cyclone over the

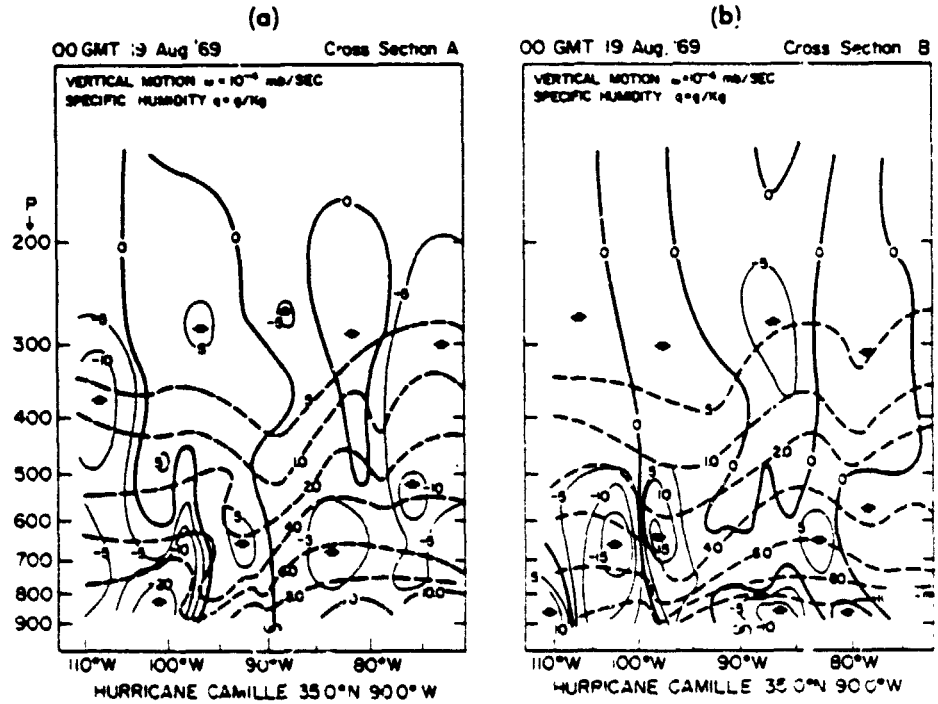
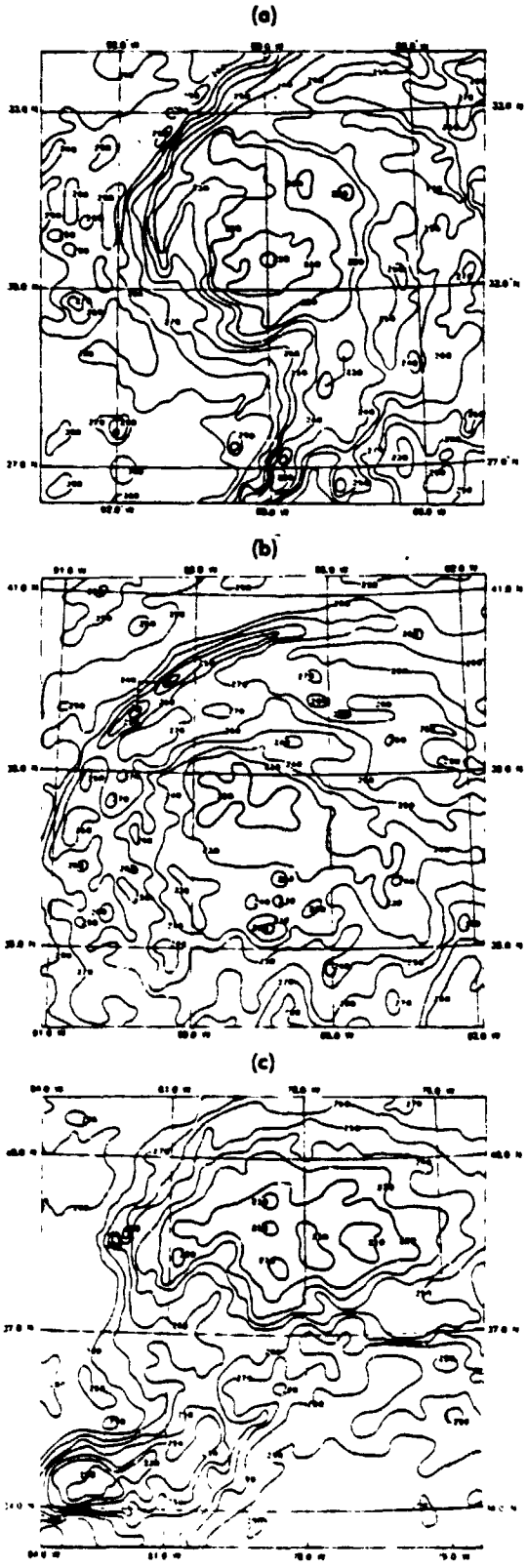


FIG. 20. As in Fig. 19 except for 0000 GMT 19 August 1969



ocean with conventional data (Shenk, 1971). In the case of Camille, the vertical cloud growth to high levels was noticeable when the heavy rains were in progress. Since the HRIR observation interval is 24 h, it is not possible to determine whether or not the clouds reached the observed 0600 GMT 20 August levels before or at the onset of the period of maximum rainfall rate. If infrared measurements had been available on a nearly continuous basis from a geosynchronous satellite, it is possible that flood warnings could have been issued based on the change in the absolute levels reached by the cloud tops. Both the Washington and Pittsburgh radars observed the strong echoes at ranges of generally $\geqslant 200$ km. Schwarz (1971) concluded that the measurements did indicate the presence of the important vertical motions. However, if the rainfall activity had been outside of effective radar range, then most likely the satellite measurements would have been the only source of information with a timely observation interval.

4. Conclusions

Prior to the rapid deepening phase of Camille, there was the development of strong low-level inflow of high moisture content air from the region near the ITCZ moving over a sea surface with temperatures 1–3 standard deviations above normal. The concurrent development of large-scale cloud bands southeast of the center, extending to the cirrus level, was the satellite evidence of the inflow. Just before the onset of the rapid deepening, the cloud bands showed a broad connection to the cloudiness within the ITCZ, and the 6.7 μm water vapor channel indicated the outflow of cirrus to the north of the center, which was associated with the necessary upper tropospheric ventilation of air away from the cyclone. It was possible to ascertain the relative strength and position of the important features, such as upper tropospheric shear lines and easterly waves, with the combination of Nimbus 3 MRIR channels and continuity.

During the rapid deepening phase, the cloud bands southeast of the center persisted. At the beginning of the rapid deepening, the size of the intense cloud region associated with Camille sharply contracted and then expanded again as the maximum deepening rate occurred. Simultaneously, a large area of subsidence developed west and north of the center. The subsidence extended from the upper troposphere vertically downward, to at least the low cloud level, as inferred from the measurements made by four of the MRIR channels. Neither the expanded cloud shield nor the large subsidence region seemed to persist for more than 24 h, leading to the hypothesis that the rapid deepening was

FIG. 21. Tropical depression Camille at (a) 0540 GMT 18 August, (b) 0450 GMT 19 August and (c) 0550 GMT 20 August 1969, as depicted by 3.8 μm T_{BB}'s from the Nimbus 3 HRIR. The isotherms are given in kelvins.

associated with a sudden surge that extended past a sustainable energy level. There was a 2–3 km reduction in maximum cloud-top level about 12 h after Camille moved inland, in response to the probable weakening of the vertical circulation, within the inner portion of the cyclone. The outflow acted as a partial obstruction to the environmental flow in the upper troposphere northwest of the center of Camille. This interaction produced horizontal convergence and subsidence. The effects of the subsidence could be seen in the two water vapor channels, where the subsidence areas were demarcated by T_{BB}'s that were initially relatively high and were increasing with time. A smaller region of dry air, produced by a shear line east of Camille, was not seen as clearly in the satellite measurements as the area northwest of the center due to the cirrus clouds in the outflow region overriding the dry air.

As Camille turned eastward across the lower Ohio valley, the cloud pattern looked like that of a small, vigorous, extratropical storm with maximum cloud tops at 13–14 km, and the interaction between the outflow and the environmental flow had diminished considerably. Speed and direction divergence in the upper tropospheric wind field, determined from conventional and cloud motion measurements, was an indication that conditions were favorable for renewed vertical cloud development as early as 1200 GMT 19 August. When the excessive rainfall was in progress over Virginia, the cloud tops had risen to 15–16 km, which was the level reached when the cyclone was still a severe hurricane just after crossing the Mississippi coast.

Since 1969 there have been many advances in satellite measurements and techniques. Latent heat estimates are now possible from 19 GHz microwave scanners (Adler and Rodgers, 1977). From geosynchronous altitude SMS/GOES measurements are being used to compute cloud motion derived wind fields that are superior to those that could be made from ATS 1 or 3 and to continuously estimate cloud-top heights from the inboard channel (Gentry *et al.*, 1976). Some of the concepts discussed in this paper are being evaluated using the SMS/GOES measurements while others can be once water vapor channel measurements are made from geosynchronous orbit starting with the European satellite which is currently scheduled for launch in November 1977.

Acknowledgments. The authors wish to thank Professor T. Theodore Fujita, University of Chicago, who computed the cloud motions, performed streamline analyses from the results, and made numerous helpful suggestions.

REFERENCES

Adler, R. F., and E. B. Rodgers, 1977: Satellite-observed latent heat release in a tropical cyclone. *Mon. Wea. Rev.*, 105, 956–963.

Barr, S., P. E. Long and I. A. Miller, 1970: Atmospheric sensing and prediction projection project. Sci. Rep. No. 2, Air Force Cambridge Research Laboratories, Bedford, Mass.

Browner, S. P., W. L. Woodley and C. G. Griffith, 1977: Diurnal oscillation of the area of cloudiness associated with tropical storms. *Mon. Wea. Rev.*, 105, 856–864.

Gentry, C. R., E. B. Rodgers, W. E. Shenk and V. Oliver, 1976: Deriving winds for hurricanes using short interval satellite imagery. *Preprints Seventh Conf. Aerospace and Aeronautical Meteorology and Symp. on Remote Sensing from Satellites*, Melbourne, Fla., Amer. Meteor. Soc., 115–118.

Hasler, A. F., W. Shenk and W. Skillman, 1976: Wind estimates from cloud motions: Phase I of an *in situ* aircraft verification experiment. *J. Appl. Meteor.*, 15, 10–15.

Hubert, L. F., and L. F. Whitney, 1971: Wind estimation from geostationary-satellite pictures. *Mon. Wea. Rev.*, 99, 665–672.

Nimbus 3 Users Guide, 1969: Nimbus Project, National Space Science Data Center, Goddard Space Flight Center, Greenbelt, Md., 237 pp.

Rodgers, E. B., V. V. Salomonson and L. Kyle, 1976: Upper tropospheric dynamics as reflected in Nimbus 4 THIR 6.7 μm data. *J. Geophys. Res.*, 81, 5749–5758.

Schwarz, F. K., 1971: The unprecedented rains in Virginia associated with the remnants of Hurricane Camille. *Mon. Wea. Rev.*, 98, 851–859.

Shenk, W. E., 1971: Meteorological satellite infrared views of cloud growth associated with the development of secondary cyclones. *Mon. Wea. Rev.*, 98, 861–868.

——, R. J. Holub and R. A. Neff, 1976: A multispectral cloud type identification method developed for tropical ocean areas with Nimbus 3 MRIR measurements. *Mon. Wea. Rev.*, 104, 284–29[illegible].

——, H. Powell, V. V. Salomonson and W. R. Bandeen, 1971: Meteorological uses of the stereographic horizon map projection. *J. Appl. Meteor.*, 10, 582–589.

Simpson, R. H., A. L. Sugg and Staff, 1970: Atlantic hurricane season of 1969. *Mon. Wea. Rev.*, 98, 307–314.

U. S. Naval Oceanographic Office, 1967. *Oceanographic Atlas of the North Atlantic Ocean*. Pub. No. 700, Washington, D. C.

Paper 55

Reprinted from Preprint Volume: Eleventh Conference on Severe Local Storms, October 2-5, 1979 Kansas City, Missouri. Published by the American Meteorological Society, Boston, Mass.

THE ROLE OF UPPER TROPOSPHERIC JET STREAKS AND LEE SIDE CYCLOGENESIS IN THE DEVELOPMENT OF LOW LEVEL JETS IN THE GREAT PLAINS

Louis W. Uccellini

Goddard Laboratory for Atmospheric Sciences
Code 914
NASA/Goddard Space Flight Center
Greenbelt, MD 20771

1. INTRODUCTION

The interaction between upper and lower tropospheric jet streaks is widely recognized as an important factor in the development of organized convective storm systems (see Newton, 1967; Danielsen, 1974; among others). The tendency has been to treat the low level (boundary layer) jet and upper tropospheric jet streak as separate entities. However, in a recent paper by Uccellini and Johnson (1979), an extensive analysis of the 10-11 May 1973 tornado outbreak in Ohio revealed that, in this particular case, the jets were not separate entities but instead represented a coupled phenomenon. Results from the 10-11 May case study showed that 1) a low-level jet (LLJ) beneath the exit region of an upper tropospheric jet streak was embedded in the lower branch of an indirect circulation, 2) intensification of the lower branch and development of the LLJ was largely a result of an increased isallobaric wind component and 3) the development of the low-level jet was coupled to the upper tropospheric jet streak by the two-layer mass adjustment within the exit region of the streak. The isallobaric wind component was the primary reason for the axis of the LLJ being at a significant angle to the upper jet's axis and the resulting veering of the wind with height. In the exit region, the geometry of this adjustment combined with the warm, moist, lower tropospheric air to the right of the jet streak and cool, dry air at the jet streak level produced the differential advections that convectively destabilized the atmosphere.

The purpose of this paper is to address the problem of applying the concept of coupled jet streaks to the large number of low-level jets which occur in the Great Plains (Bonner, 1968) and which are known to be important for the development of convective storm systems (Means, 1952, 1954; Pitchford and London, 1962; Bonner, 1966). In Section 2, a literature review is presented which notes that boundary layer and terrain effects are emphasized as causative factors in the development of low level jets in the Great Plains, while upper tropospheric characteristics and processes have not been considered. In Section 3, 15 cases of LLJ's in the Great Plains that have been previously discussed in the literature are reviewed with the emphasis placed on determining if any systematic synoptic forcing is common to these cases. A summary of results is presented in Section 4.

2. LOW LEVEL JETS IN THE GREAT PLAINS

Bonner's (1968) statistical analysis confirmed that a large number of low level jets occur in the Great Plains, with the maximum number of LLJ's existing from Texas to Nebraska (Fig. 1). Means (1952, 1954) and Bonner (1966) have shown that these low level jets are especially important for their rapid transport of heat and moisture from the Gulf region into areas of convective storms which produce heavy rainfall. The LLJ's in the Great Plains region are characterized by a diurnal oscillation, as the wind speeds reach maximum intensity by early morning, and are associated with the development of a nocturnal temperature inversion (Blackadar, 1957; Wexler, 1961; Hoecker, 1963; Izumi and Barad, 1963; Izumi, 1964; Lettau, 1967, Bonner, 1968). The westward extension of the North Atlantic Subtropical High (Wexler, 1961), boundary layer mixing processes (Blackadar, 1957), and the diurnal radiation cycle over sloped terrain (Lettau, 1967), with greater emphasis placed upon the topographical characteristics by Paegle and Rasch (1973) and Paegle (1978), have all been related to the generation of the LLJ and its seasonal, temporal and geographic preference.

Reiter (1969), Danielsen (1974), Newton (1956, 1967) all present evidence that low level jets can also develop in response to synoptic or subsynoptic scale processes particularly through a response to lee side cyclogenesis common to the Great Plains (see Hovanec and Horn, 1975). For the large number of jet cases collected for the climatological summary, Bonner (1963) stated that

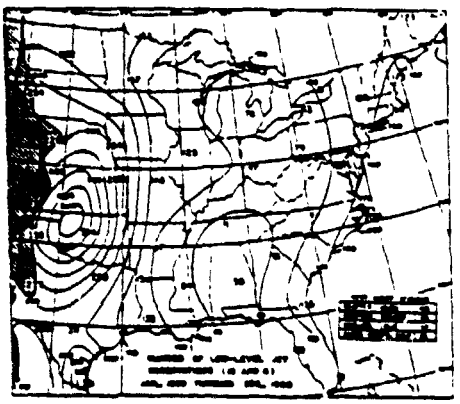


Fig. 1: Number of low level jet observations from January 1959 through December 1960 at 1200 GMT and 0000 GMT. (From Bonner, 1965.)

TABLE 1
Cases of Low Level Jets

Date	Author	Type
14 July 1959*	Bonner (1965)	2
14 August 1959*		1
20 August 1959*		1
19 April 1960		1
22 April 1960		1
23 April 1960		1
10 July 1960		1
23 August 1960		1
2 December 1960		1
15 March 1961	Izumi (1964)	2
23 April 1961	Hoecker (1963)	1
28-29 May 1961		2
30 May 1961		1
16-17 May 1961	Bonner (1963, 1966)	1
17-19 November 1948**	Newton (1956)	1

Type 1: Trough upstream and ridge downstream of southern Great Plains with 300 mb jet streak propagating into region.

Type 2: Ridge located directly over Great Plains with weak upper tropospheric winds.

*250 mb charts reviewed for upper level analysis.
**actual winds on 300 mb charts not available.

"On roughly 60 percent of the jet days at each station, cold fronts or low pressure centers were to be found within 350n mi to the west of the station. On roughly one-half of these days, frontal passage occurred within the next twelve hours." In a recent discussion, Bonner[1] stated that the organized, coherent LLJ's in the Great Plains that could be analyzed within a region (rather than being obvious at only a few individual stations) were frequently associated with lee side troughing, lee side cyclogenesis or a frontal passage associated with a cyclone further north. These observations suggest a high correlation between LLJ occurrence and synoptic to subsynoptic scale forcing. They also serve as a motivation for reviewing cases of LLJ's previously reported on in the literature for which boundary layer processes and terrain effects were emphasized and upper tropospheric, synoptic features were not considered.

3. REVIEW OF LLJ CASE STUDIES

An extensive amount of research into the forcing of LLJ's was undertaken in the 1960's with special PIBAL networks and tower measurements essentially to test the theories previously presented by Blackadar (1957) and Wexler (1961). Table 1 lists 15 cases of LLJ's which were used in these studies and include 4 cases from 1961 (Hoecker 1963; Bonner 1963, 1966) for which special network data are available. The cases listed in Table 1 span all the seasons and include situations with and without convective storms. Newton's (1956) study of lee side cyclogenesis is listed in Table 1 since it also included a description of a strong low level jet in the Great Plains. Except for the Newton paper, none of the case studies in Table 1 include any meteorological charts or other information above the 700 mb level.

As a first step in reviewing the previously documented cases of the low level jet, upper air maps were collected for each case, reviewed and catagorized as several basic flow

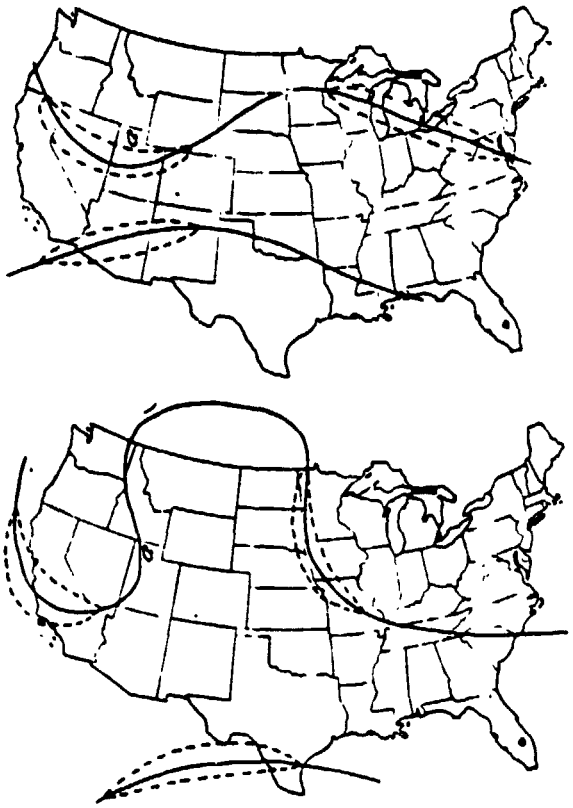


Fig. 2: Schematic of upper tropospheric (300 mb) flow patterns for 15 cases of LLJ's in southern Great Plains. Top: type 1 condition which is representative for 12 out of the 15 cases. Bottom: type 2 condition which existed for the other 3 cases. Dashed lines are jet streaks.

patterns became readily apparent. The schematic in Fig. 2 summarizes the upper tropos heric flow patterns which prevailed during the occurrence of the LLJ and shows that two basic patterns existed for these cases. The first type consists of a trough over the Rockies and ridge located in the eastern third of the country with significant upper tropospheric jet streaks propagating toward the Great Plains from the Nevada-California region (polar origin) and from the Arizona-Mexico region (subtropical origin). These conditions existed for 12 out of the 15 cases. There is considerable variability in the magnitude of the trough and upper tropospheric jet streaks located over the western United States for the 12 LLJ cases listed in Table 1. However, the existence of a 300 mb trough over the far west, upper tropospheric jet streaks propagating toward the Great Plains and the development of a lee side cyclone or trough that occurs with this type of upper tropospheric flow (Newton, 1956; Hovanec and Horn, 1975) is remarkably consistent. The second pattern, which existed for 3 out of 15 cases, consists of a strong ridge located over the front range of the Rockies with weak upper tropospheric flow over the north-Texas, western Oklahoma-Kansas region. The well documented LLJ cases which clearly display the diurnal wind oscillation with a nocturnal maximum coinciding with a boundary layer inversion (e.g. Izumi 1964; Hoecker 1963) were associated with

[1]Personal Communication

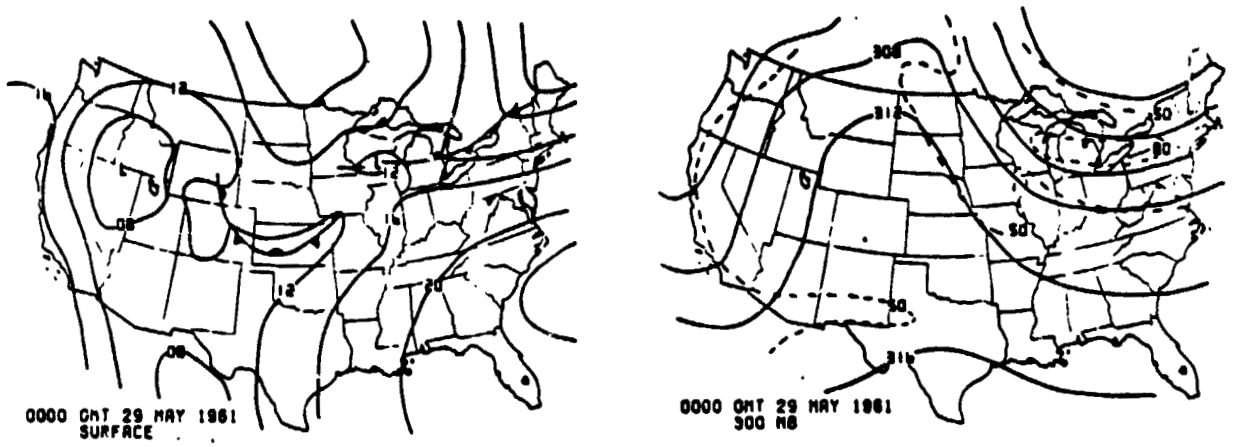


Fig. 3: (A) surface and (B) 300 mb National Weather Service analyses for 0000 GMT 29 May 1961. Surface isobars in mb (12=1012 mb). 300 mb heights in geopotential feet (308=30,800 feet) and isotachs in knots.

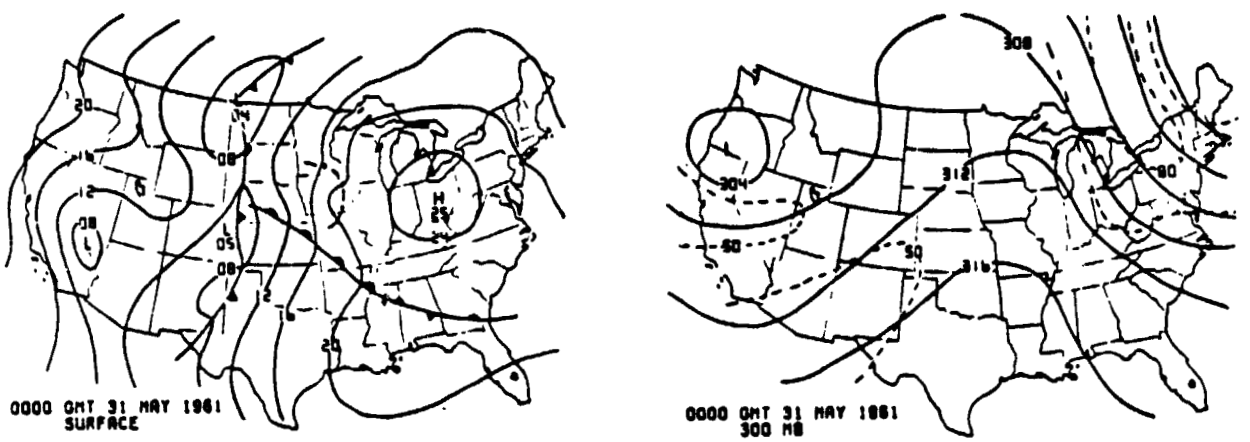


Fig. 4: Same as 3 except for 0000 GMT 31 May 1961.

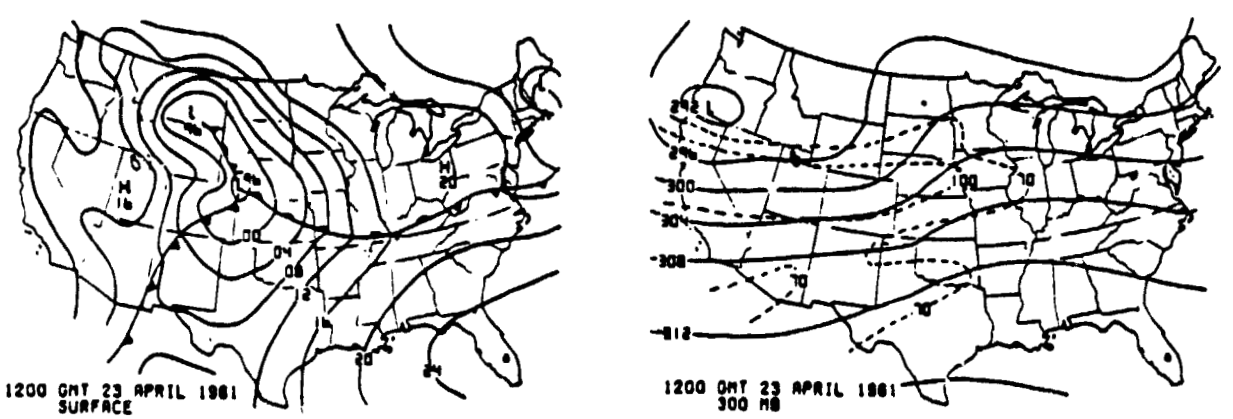


Fig. 5: Same as 3 except for 1200 GMT 23 April 1961.

this type of flow.

The three Hoecker cases (Table 1) illustrate the variable nature of the LLJ's observed in the southern Great Plains during the special observation period in 1961 and also provide evidence on the relative influence that boundary layer processes have on the LLJ as a function of the synoptic scale forcing. The 28-29 May 1961 case of a LLJ in the southern Great Plains illustrates the "classic" diurnal oscillation in the magnitude and coherency of the LLJ (see Fig. 5 in the Hoecker paper). The LLJ, which remained just above the boundary layer temperature inversion (400 m), reached a maximum value of 25 m s^{-1} between 0000-0600 CST, weakened immediately after sunrise to a 15 m s^{-1} minimum and increased again after sunset. The LLJ appeared well organized during the night but appeared to be less coherent during the day, apparently as a result of the solar insolation and increased boundary layer turbulence. The surface maps for 28 May display a relatively weak pressure gradient in the southern Great Plains associated with a weakening inverted trough in Oklahoma (Fig. 3A). The 300 mb flow is also weak in the southern Great Plains with the height contours illustrating a type 2 condition defined in Table 1 (Fig. 3B).

The 30-31 May 1961 case from Hoecker provides additional evidence of a diurnal oscillation but also shows a deviation from the classic pattern. During the early morning of May 30, the LLJ increased to 20 m s^{-1} over Oklahoma and remained at the 400 m level coinciding with the inversion level (see Fig. 7 in the Hoecker paper). Immediately after sunrise, the LLJ appeared to break down, but more so in the coherency rather than in the magnitude of maximum velocity which dropped off to 15 m s^{-1}. The LLJ began reorganizing and increasing in magnitude during the afternoon rather than after sunset and increased to 25 m s^{-1} by 2200 CST 30 May. The surface map for 30 May shows a developing pressure gradient associated with a lee side trough (Fig. 4A) as the 300 mb trough shifted east from its 28 May position (Fig. 4B). Relatively weak jet streaks propagated toward the southern Great Plains with the exit region of the subtropical jet coinciding with the position of the LLJ in the Oklahoma region.

The 23 April 1961 case from Hoecker is characterized by much larger synoptic scale forcing than the previous two cases as a major lee side cyclone developed within the exit region of a jet streak propagating toward the Great Plains from the Pacific Coast (Fig. 5). The surface pressure gradient in this case was nearly 33% larger than the other two cases. Although the 25 m s^{-1} magnitude of the LLJ observed in Oklahoma on 23 April (see Fig. 2 in Hoecker paper) was no larger than the magnitudes observed in the other cases, the persistence and general characteristics of the LLJ were noticeably different. The LLJ increased during the night of 22 April and morning of 23 April as the pressure gradient also increased in the Great Plains region in response to the lee side cyclogenesis. Unlike the other cases, the LLJ did not rapidly weaken during the morning but persisted and remained coherent well into the afternoon with the magnitude of the LLJ remaining greater than 20 m s^{-1}. Although Hoecker attributed the behavior of the LLJ in the 23 April case to a westward extension of the subtropical high and daytime cloud cover, it appears more likely that the cyclogenesis and the upper tropospheric jet streaks which are important for lee side development (Newton 1956 Hovanec and Horn, 1975) could also be responsible for the strong pressure gradient in the Great Plains and persistent nature of the LLJ for this case.

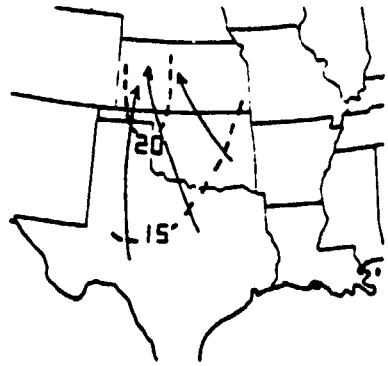
The 16-17 May 1961 case, previously analyzed by Bonner (1963, 1966) using the special PIBAL network offers additional evidence that the combined effects of upper tropospheric jet streaks and lee side cyclogenesis can influence the behavior of LLJ's in the southern Great Plains. In this case, the LLJ was well established in southwest Kansas in the afternoon of 16 May and shifted southeastward to Oklahoma by early evening (0000 GMT 17 May; Fig. 6). During the night, the magnitude of the wind maximum increased to over 30 m s^{-1} as the position of the LLJ continued to shift eastward then northeastward to southwest Missouri by 1200 GMT 17 May.

Fig. 6 also includes the surface pressure tendencies computed over a two hour interval by Bonner (1963) and smoothed to eliminate high frequency perturbations related to individual thunderstorm cells. Bonner's study attempted to relate the surface pressure tendencies and resultant isallobaric wind to the evolution of the LLJ. Bonner noted that at any given time the isallobaric, ageostrophic wind tended to be perpendicular to the axis of the LLJ and that the geostrophic wind was a better approximation to the real wind.

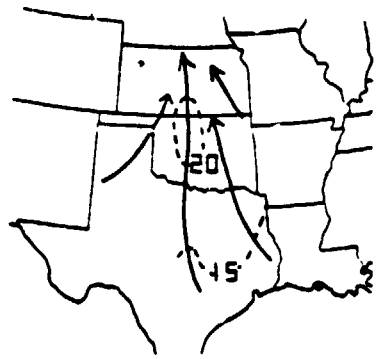
However, it appears from Fig. 6 that the surface pressure tendencies and the eastward shift of the confluence zone could both continuously contribute to parcel accelerations into the observed locations of the LLJ. Between 1800 GMT and 0000 GMT the area of maximum pressure falls shifts southeastward from the Texas panhandle to north-central Texas and then by 0600 GMT northeastward to southwest Missouri. The relative positions of the confluence zone and the pressure falls area upwind of the LLJ core would both contribute to parcel accelerations in the "along-stream" direction into the core of the LLJ. For example, lower tropospheric parcels located in northeast Texas at 0000 GMT would have an ageostrophic component directed to the west approximated by the confluence term (2) and the isallobaric term (1) in the equation:

$$\underset{\sim}{V}_{ag} = \tfrac{1}{f}\left(\underset{\sim}{k} \times \frac{\partial \underset{\sim}{V}}{\partial t}\right) + \tfrac{1}{f}\left(\underset{\sim}{k} \times \underset{\sim}{V} \cdot \nabla \underset{\sim}{V}\right), \quad (1)$$

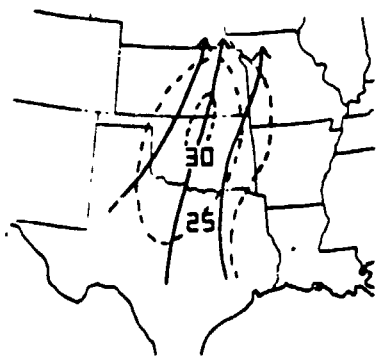
where $\underset{\sim}{V}_{ag}$ is the ageostrophic wind, $\underset{\sim}{V}$ is horizontal wind, ∇ gradient operator and $\underset{\sim}{k}$ the unit vector normal to the horizontal plane. The ageostrophic component ($\underset{\sim}{V}_{ag}$) would, in turn, lead to parcel accelerations toward the north-central Oklahoma where the LLJ was located by 0600 GMT. Similarly given the confluence zone in north Texas and the axis of negative $\partial p/\partial t$ from northeast Texas to northeast Kansas at 0600 GMT, parcels located in northeast Oklahoma down to Texas at 0600 GMT would undergo accelerations toward southwest Missouri where the LLJ was located at 1200 GMT 17 May. The evolution of the wind, height and pressure tendency fields in this fashion is consistent with a mutual and continual mass-momentum adjustment occurring on a subsynoptic scale in association with the eastward propagation of the cyclone.



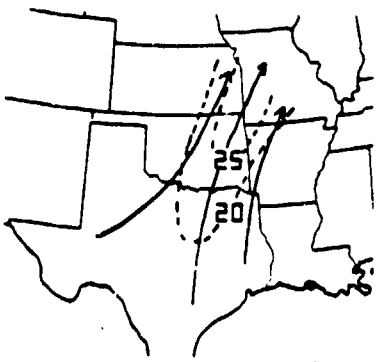
1200 CST 16 MAY 1961
1800 GMT 16 MAY 1961



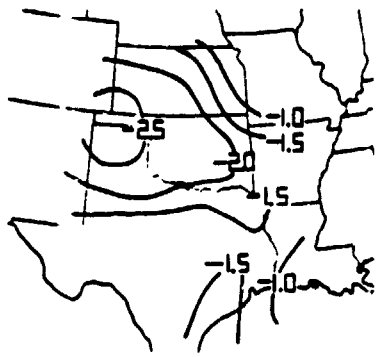
1800 CST 16 MAY 1961
0000 GMT 17 MAY 1961



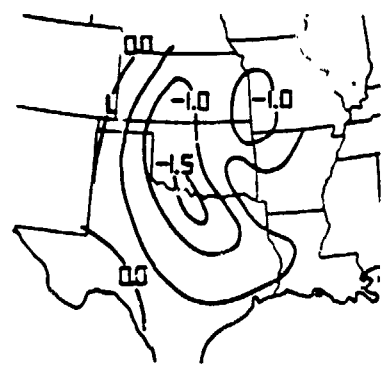
0000 CST 17 MAY 1961
0600 GMT 17 MAY 1961



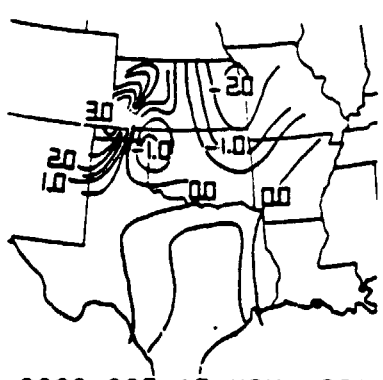
0600 CST 17 MAY 1961
1200 GMT 17 MAY 1961



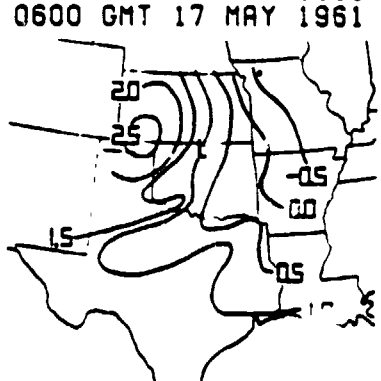
1200 CST 16 MAY 1961
1800 GMT 16 MAY 1961



1800 CST 16 MAY 1961
0000 GMT 17 MAY 1961



0000 CST 17 MAY 1961
0600 GMT 17 MAY 1961



0600 CST 17 MAY 1961
1200 GMT 17 MAY 1961

Fig. 6: Isotach (m s^{-1}) and streamline analyses at 1 km (left) and surface pressure tendencies (mb/2h) for 16-17 May 1961 (from Bonner, 1961).

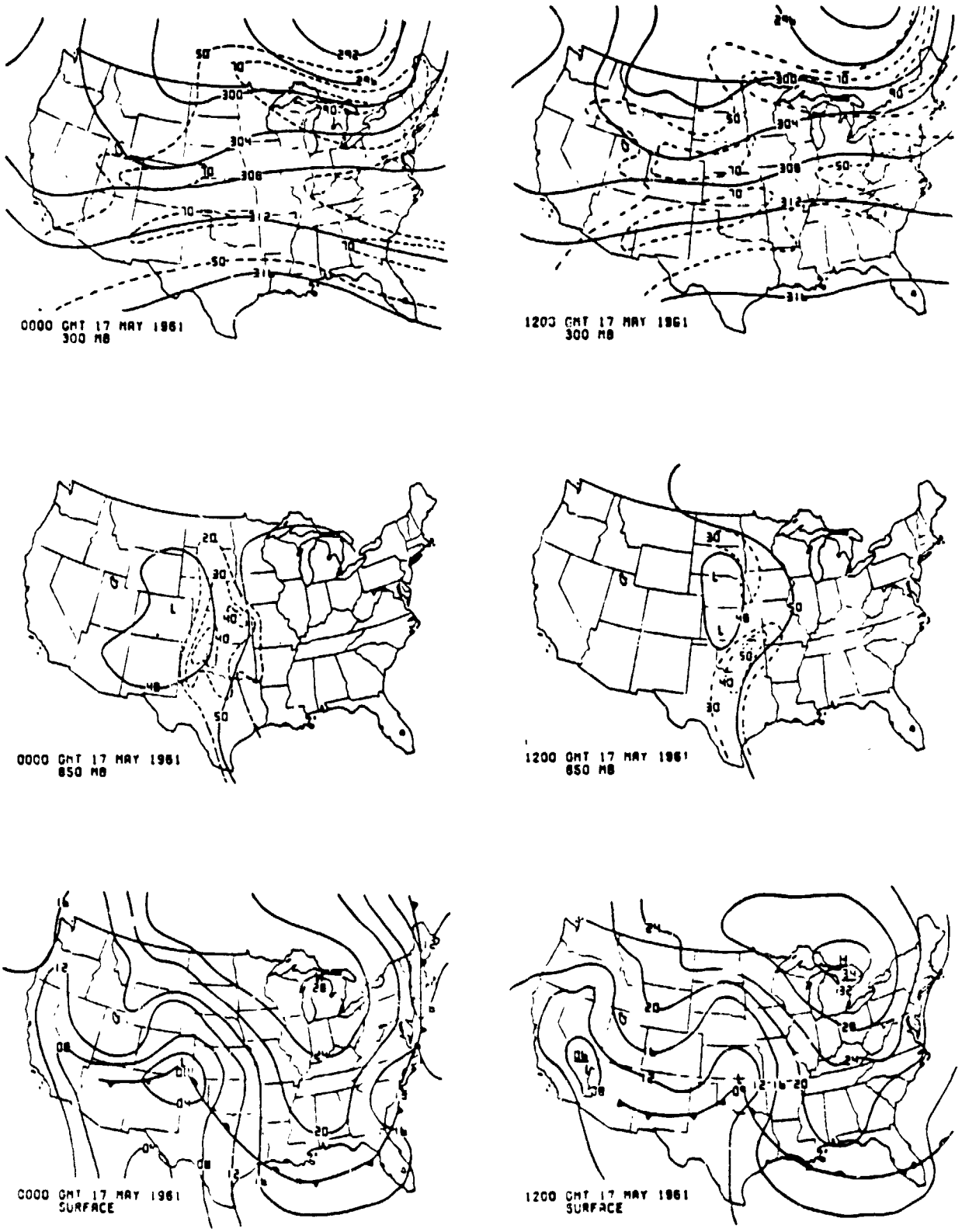


Fig. 7: 300 mb, 850 mb, surface analyses for 0000 GMT 17 May (left) and 1200 GMT 17 (right) 1961. See Fig. 3 for details. Isotachs for 850 mb and 300 mb analyses are in knots.

The upper tropospheric features and synoptic scale characteristics of the 16-17 May 1961 case are illustrated in Fig. 7. Between 0000 GMT and 1200 GMT, two upper tropospheric jet streaks propagated eastward into the Great Plains region. Lee side cyclogenesis terminated by 0000 GMT 17 May as the surface low filled by 1200 GMT (note Fig. 6 that positive pressure tendencies were located in the center of the low by 1200 GMT). The 850 mb maps in Fig. 7 and the isotach maps at 0000 GMT and 1200 GMT 17 May in Fig. 6 reveals that at both times the LLJ was located in the exit region of the southern upper level jet streak. Combining the information from Figs. 6 and 7 suggests that the evolution of the LLJ, especially its eastward shift during the 12 h period, is linked to the upper level jet's propagation and associated mass adjustments as discussed by Uccellini and Johnson (1979). It would of course take a thorough analysis to confirm this interpretation and to determine the relative importance of these processes as compared to the boundary layer processes which could also contribute to the increase in wind speed observed between 0000 GMT and 0600 GMT (Fig. 6).

4. SUMMARY

In a recent paper, Uccellini and Johnson (1979) discussed the manner in which the development of LLJ's could be coupled to upper tropospheric jets through mutual mass-momentum adjustments within the exit region of the upper level jet streak. In this paper, 15 cases of LLJ's that were previously discussed in the literature were reviewed to determine if the coupling concept has any relevance to the large number of LLJ's observed in the southern Great Plains. The literature review revealed that most research efforts have emphasized boundary layer processes and terrain effects in an effort to explain the development of the LLJ and its geographic and temporal preferences. The review also revealed an almost total lack of information concerning the upper tropospheric conditions for these cases.

In 12 out of the 15 cases, the synoptic pattern was characterized by upper tropospheric jet streaks propagating toward the Great Plains from the Rocky Mountain area with the surface pressure gradients increased by lee side cyclogenesis or lee side troughing (type 1). In these cases the LLJ's were located within the exit region of the upper level jet and directed toward the cyclonic side. In the other 3 cases, the upper troposphere over the Great Plains was characterized by a significant ridge and weak upper tropospheric flow (type 2). A more detailed review of the Hoecker (1963) and Bonner (1963, 1966) cases indicate that the nature of the LLJ seems dependent upon the magnitude of the synoptic scale forcing. The cases of well documented LLJ's characterized by the classic diurnal oscillation were associated with the type 2 upper tropospheric pattern and relatively weak surface pressure gradients. The cases which had significant upper level jet streaks propagating toward the Great Plains and lee side cyclogenesis had LLJ's which deviated from the classic pattern. In these cases, the LLJ was well defined, coherent and more persistent even in the afternoon and extended above the planetary boundary layer. However, there was still a tendency for the maximum winds to be observed in the early morning suggesting that,

Fig. 8: Sites of cyclogeneses-spring (top) and fall (bottom). Solid circles or triangles indicate cyclones which originated west of the continental divide but intensified in the Colorado cyclogenetic area. (From Hovanec and Horn, 1975).

even with significant synoptic scale forcing, boundary layer and terrain effects can still increase the magnitude of the LLJ in this region. Finally, the review of Bonner's case (1963, 1966) indicated that the evolution of the LLJ seems to be coupled to the propagation of an upper tropospheric jet streak and weakening lee side cyclone. The LLJ was located within the exit region of an upper level jet streak at two successive radiosonde observing periods. The LLJ also seemed to respond to an evolving surface pressure tendency field in a manner consistent with mutual mass-momentum adjustment concepts.

It is suggested that the subsynoptic forcing associated with the upper tropospheric jet streak's role in lee side cyclogenesis, as discussed by Newton (1956) and Hovanec and Horn (1975), are important in the forcing of the LLJ's in the southern Great Plains. While the importance of boundary layer and terrain effects in forcing a diurnal oscillation of the LLJ is evident, other factors besides boundary processes should also be considered to explain the large number and evolution of the LLJ's observed in the southern Great Plains. One factor that has to be questioned is the concept that the westward extension, or retrogression, of the North Atlantic subtropical high, creates the pressure gradient force needed for the development of the LLJ. At least for these cases of LLJ's, it appears that the high pressure cell located in the southeast United States is of polar origin and that the pressure gradients increase over the Great Plains in response to a developing low pressure system to the west of the region. Given this type of synoptic to subsynoptic scale forcing in the Great Plains region, one must also question the assumption of imposing a constant pressure gradient or a periodic variation in the pressure gradient for studying the total evolution of LLJ's in the Great

Plains.

The questions raised by this review can basically be summarized by comparing the climatological summaries of LLJ's by Bonner (Fig. 1) and lee side cyclogenesis by Hovanec and Horn (Fig. 8). The coincidence of maxima for both phenomena suggests a significant correlation between the two. Detailed statistical analysis being completed by Horn, Achtor and Hovanec (1979) and additional detailed case studies are needed to prove that this correlation is significant and that subsynoptic scale processes associated with upper level jet streaks and lee side cyclogenesis are indeed an important forcing mechanism for the development of LLJ's in the Great Plains.

5. REFERENCES

Blackadar, A. K., 1957: Boundary layer wind maxima and their significance for the growth of nocturnal inversions. Bull. Amer. Meteor. Soc., 38, 283-290.

Bonner, W. D., 1961: An experiment in the determination of geostrophic isallobaric winds from NSSP pressure data. Res. Paper, 10. Thunderstorm Project, Univ. of Chicago, 26 pp.

_____, 1965: Statistical and kinematic properties of the low-level jet stream. Res. Paper 38. Satellite and Mesomet. Res. Proj., Univ. of Chicago, 34 pp.

_____, 1966: Case study of thunderstorm activity in relation to the low-level jet. Mon. Wea. Rev., 94, 167-178.

_____, 1968. Climatology of the low level jet Mon. Wea. Rev., 96, 833-850.

Danielson, E. F., 1974: The relationship between severe weather, major dust storms and rapid cyclogenesis. Synoptic Extratropical Weather Systems. M. Shapiro, Ed., National Center for Atmospheric Research, 213-241.

Beecher, W. H., 1961: Three southerly low-level jet streams delineated by the Weather Bureau special pibal network of 1961. Mon. Wea. Rev., 91, 573-582.

Horn, L. H., T. H. Achtor, and R. D Hovanec, 1979: The mean upper tropospheric jet stream, associated low level jet and static stability for spring season Colorado cyclones. To be presented at 11th Conference for Severe Local Storms.

Hovanec, R. D. and L. H. Horn, 1979: Static stability and 300 mb isotach field in the Colorado cyclogenetic area. Mon. Wea. Rev., 103, 630-638.

Izumi, Y., 1964: The evolution of temperature and velocity profiles during breakdown of a nocturnal inversion and a low-level jet. J. Appl. Meteor., 3, 70-82.

_____, and M. L. Barad, 1963: Wind and temperature variations during development of a low-level jet. J. Appl. Meteor., 2, 668-673.

Lettau, H. H., 1967: Small to large scale features of boundary structures over mountain slopes. Proc. Symp. Mountain Meteor., Colorado State University, 1-74.

Means, L. L., 1952: On thunderstorm forecasting in the central United States. Mon. Wea. Rev., 80, 165-189.

_____, 1954. A study of the mean southerly wind maxima in low levels associated with a period of summer precipitation in the middle west. Bull. Amer. Meteor. Soc., 35, 166-170.

Newton, C. W., 1956: Mechanisms of circulation change during a lee cyclogenesis. J. Meteor., 13, 528-539.

_____, 1967: Severe convective storms. Vol. 12, Advances in Geophysics. Academic Press, 257-308.

Pagnia, J., 1978. A linearized analysis of diurnal boundary layer convergence over the topography of the United States. Mon. Wea. Rev., 106, 492-502.

Pagnia, J., and C. E Kaseh, 1973 Three-dimensional characteristics of diurnally varying boundary-layer flows. Mon. Wea. Rev., 101, 746-756.

Pitchford, K., and J. London, 1962: The low-level jet as related to nocturnal thunderstorms over Midwestern United States. J. Appl. Meteor., 1, 43-47

Reiter, E. R., 1969 Tropopause circulation and jet streams. World Survey of Climatology, Vol. 4. Climate of the Free Atmosphere, Lam, Ed., 85-193.

Uccellini, L. W., and D. R. Johnson, 1979: The coupling of upper and lower tropospheric jet streaks and implications for the development of severe convective storms. Mon. Wea. Rev. June issue.

Wexler, H., 1961. A boundary layer interpretation of the low-level jet. Tellus, 13, 368-378.

Paper 56

Reprinted from *Monthly Weather Review*, 107, pp. 682-703, 1979.

The Coupling of Upper and Lower Tropospheric Jet Streaks and Implications for the Development of Severe Convective Storms

LOUIS W. UCCELLINI[1]

Space Science and Engineering Center, University of Wisconsin, Madison 53706

DONALD R. JOHNSON

Space Science and Engineering Center and Department of Meteorology, University of Wisconsin, Madison 53706

(Manuscript received 17 July 1978, in final form 2 March 1979)

ABSTRACT

Transverse circulations in the exit and entrance regions of jet streaks are investigated through numerical simulation, a case study, and an application of the isallobaric wind equation in isentropic coordinates, to study the interaction between upper and lower tropospheric jets and the development of severe convective storms. A hybrid isentropic-sigma coordinate numerical model is used to simulate the mass and momentum adjustments associated with a jet streak propagating in a zonal channel. The numerical results depict a two-layer mass adjustment in the exit and entrance region of the jet streak. The results also verify that the isallobaric wind on lower isentropic surfaces is a primary component of the return branches of transverse circulations and is forced by the two-layer mass adjustment accompanying the propagating jet streak. Results from the case study of a severe weather outbreak show that 1) a low-level jet (LLJ) beneath the exit region of an upper tropospheric jet streak is embedded in the lower branch of an indirect circulation, 2) intensification of the lower branch and development of the LLJ is largely a result of an increased isallobaric wind component, and 3) the development of the LLJ is coupled to the upper tropospheric jet streak by the two-layer mass adjustment within the exit region of the streak. The isallobaric wind component of the LLJ is the primary reason for the axis of the LLJ being at a significant angle to the upper jet's axis and the resulting veering of the wind with height. In the exit region, the geometry of this adjustment, combined with warm, moist, lower tropospheric air to the right and ahead of the jet streak and cool, dry air at the jet streak level, produced the differential advections that convectively destabilized the atmosphere. Results of the case study support the concept that the development of conditions favorable for severe convective storms can be forced by mass and momentum adjustments which accompany the propagation of an upper tropospheric jet streak.

1. Introduction

The concept that an interaction between upper and lower trophospheric jet streaks is important for the development of organized convective storm systems was implied in the work of Fawbush and Miller (1953, 1954). They noted that advection of cool, dry air associated with strong westerly winds in the middle troposphere and the rapid northward movement of a tongue of warm, moist air in the low troposphere create favorable conditions for deep convection. They also emphasized that the likelihood of convection is increased with strong veering of winds, so that the angle between the mid-tropospheric wind and the axis of the moisture tongue is large. Petterssen (1956) and Newton (1967), among others, related veering of winds with height and subsequent differential moisture and temperature advections to the intersection of jet axes (Fig. 1) and the development of convective instability. The greatest convective instability develops when the jet axes tend to be orthogonal and the intensity of the advections is maximized.

The roles of tropospheric jet streaks[2] in the development of severe convective storms are summarized by Petterssen (1956), Reiter (1963), Ludlam (1963), Newton (1963, 1967), Palmén and Newton (1969) and Danielsen (1974). In these models the primary role of the upper trophospheric jet streak is to advect cool, dry air within the upper and middle troposphere, to enhance upper level divergence, and to transport the sensible heat downstream from the convective region. The low-level jet (LLJ) rapidly transports heat and moisture toward the convective region (see also Means, 1952, 1954; Bonner, 1966). The combined

[1] Present affiliation: NASA/Goddard Space Flight Center, Laboratory for Atmospheric Sciences, code 914, Greenbelt, MD 20771.

[2] Palmén and Newton (1969, p. 199) define jet streaks as the regions of isotach maxima.

0027-0644/79/060682-22$09.50

C-6

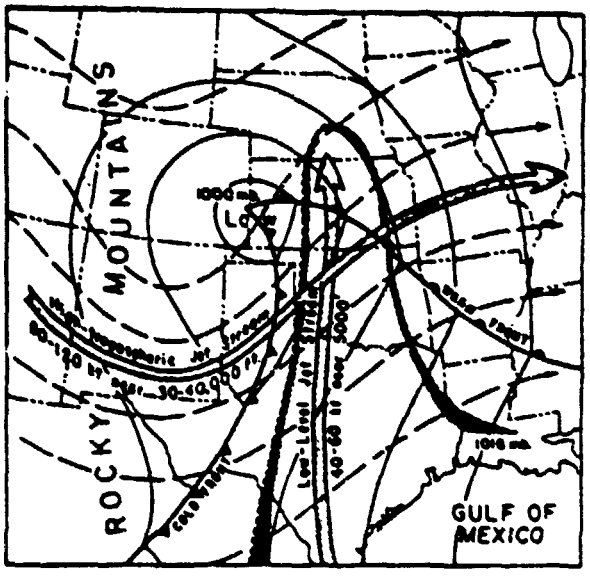


FIG. 1. Schematic features of a severe weather outbreak. Solid lines are sea level isobars; dashed lines streamlines of upper tropospheric flow. Shading outlines general area of low-level moisture tongue and region of potential instability (from Newton, 1967).

role of the upper and lower trophospheric jet streaks is to create a region of convective instability within which the severe weather ultimately occurs. Beebe and Bates (1955) also related the relative orientation and position of the two jets to a superposition of divergent and convergent fields which induces mechanical lifting and initiates the convective storms. Reiter (1963) offered additional evidence of vorticity advection, divergence and vertical exchange of momentum in his emphasis that jet streaks exert a causal mechanism in the production of severe weather, while Ludlam (1963) showed that the mean position of the jet streak is closely related to the global distribution of severe convection.

With the analyses of upper and lower tropospheric jet streaks normally depicted on pressure surfaces, these jets are normally treated as separate entities in the criteria for the development of severe convective storms. However, with jet streaks analyzed on isentropic surfaces, the two apparently distinct jets appear to merge in some cases as the LLJ extends toward the middle troposphere (Reiter, 1969). Reiter suggested that, in some cases, low-level jets are not separate entities but appear to be coupled to an upper tropospheric jet streak.

The purposes of this study are to determine how upper and lower tropospheric jet streaks are coupled and to discuss the role of this coupling in the development of severe convective storms. In Section 2, types of low-level jet streaks which affect the North American continent are discussed and the type of LLJ selected for study is specified. The basis for coupling lower and upper tropospheric jet streaks that stems from adjustment concepts is presented in Section 3. A case study is discussed in Section 4 which links mass and momentum adjustments in the exit region of an upper tropospheric jet streak to the development of a LLJ and to the differential temperature and moisture transports that create favorable conditions for severe convective storms. A summary of the results and suggestions for future research is presented in Section 5.

2. Low-level jet streaks

Bonner's (1968) climatological summary illustrated the large number of lower tropospheric jet streaks that develop in the Great Plains and Midwest, with the maximum number of LLJ's occurring from Texas to Nebraska during the spring and summer months. The low-level jets in the southern Great Plains are characterized by a diurnal oscillation reaching maximum intensity by early morning, and are associated with a nocturnal temperature inversion (Blackadar, 1957; Wexler, 1961; Gerhardt, 1962, 1963; Hoecker, 1963; Izumi and Barad, 1963; Izumi, 1964; Bonner, 1968; Lettau, 1967). The seasonal shift of the subtropical high, sloped terrain, boundary-layer processes and the diurnal radiation cycle have all been related to the generation of the low-level jet and its seasonal, temporal and geographic preferences (Blackadar, 1957; Wexler, 1961; Lettau, 1967; Paegle and Rasch, 1973).

Reiter (1969) noted that lower tropospheric wind maxima also occur in the Midwest which appear to develop in response to synoptic or subsynoptic-scale forcing. Hoecker (1963) and Bonner (1966) presented examples of LLJ's in which lee side troughing or cyclogenesis is evident just east of the Rocky Mountains. These LLJ's developed with a minimal diurnal oscillation and extended above the planetary boundary layer to the 850 mb level. For similar cases, Danielsen (1974) pointed out that moisture-laden low-level wind maxima form in response to the increasing pressure gradients associated with the developing lee side cyclone. Supporting evidence for the existence of low-level jets in conjunction with cyclone evolution is presented in Browning and Harrold's (1970) Doppler radar observations of a cold front traversing the British Isles. They determined that prefrontal horizontal moisture transport is confined to the lower troposphere and is predominantly related to the lower tropospheric wind maximum located immediately ahead of and parallel to the cold front. This study is focused on synoptic-scale adjustments associated with the LLJ located beneath the exit region of the upper tropospheric jet streak, directed from the anticyclonic (south) toward the cyclonic (north) side (Fig. 1).

3. Mass and momentum adjustments associated with jet streaks

The mutual mass and momentum adjustments associated with a jet streak and the response of the lower tropospheric winds are discussed in this section. Since the LLJ located beneath the exit region of an upper tropospheric jet streak in severe weather situations is directed toward the cyclonic side of the jet (see Fig. 1), it would appear to be embedded within the return branch of the indirect circulation. Thus, isolating the forcing of a return branch of a transverse circulation through mass adjustment associated with a propagating wind maximum should provide insight into the coupling of upper and lower tropospheric jets.

Direct and indirect transverse circulations exist in entrance and exit regions of propagating jet streaks (University of Chicago, 1947; Namias and Clapp, 1949; Bjerknes, 1951; Riehl *et al.*, 1952; Murray and Daniels, 1953; Sawyer, 1956; Newton, 1959; Staley, 1960; Reiter, 1969; Johnson, 1970; Cahir, 1971; Uccellini, 1976). The indirect circulation within the exit region consists of an upper tropospheric branch directed toward the anticyclonic side of a jet, rising and sinking motion on the cyclonic and anticyclonic sides, respectively, and a lower tropospheric return branch directed toward the cyclonic side. The direct circulation in the entrance region is completely reversed from the sense of the indirect circulation.

The requirement that direct and indirect circulations must exist for unbalanced currents implicitly stems from the work of Rossby (1938, 1949), Cahn (1945) and Sawyer (1956). In a recent study of transverse circulations and subsynoptic precipitation bands, Cahir (1971) simulated direct and indirect circulations using a two-dimensional primitive equation model mapped on a vertical plane normal to a propagating jet streak and by applying the model to actual case studies. An example from Cahir's numerical simulation of an indirect circulation (Fig. 2) depicts upper and lower transverse components of 5.8 and 4.7 m s^{-1}, respectively, and vertical motions >2.0 μb s^{-1}. In a later experiment, with the initial relative humidities increased by 30%, the added latent heat release nearly doubled the magnitude of the upward vertical motion on the cyclonic side of the jet. Cahir applied his model to numerous cases and emphasized that subsynoptic-scale precipitation bands result from the upward vertical branches of both direct and indirect circulations associated with propagating jet streaks.

The verification of the transverse circulation from observational evidence is compounded by three-dimensional wave structure with its nonlinear advection associated with alongstream variation of momentum and pressure forces, stratification of the atmosphere, the influence of terrain, curvature effects, and interactions with large-scale waves during cyclone development and long-wave amplification (Newton, 1954; Beebe and Bates, 1955; Shapiro, 1975). Some three-dimensional aspects of mutual mass and momentum adjustments and coupling between the upper and lower troposphere will be identified in this paper through the geostrophic momentum approximation within an isentropic framework,

$$\frac{d\mathbf{U}}{dt} \approx \frac{d\mathbf{U}_g}{dt} \approx \frac{\partial \mathbf{U}_g}{\partial t_\theta} + \mathbf{U}\cdot\nabla_\theta \mathbf{U}_g + \frac{d\theta}{dt}\frac{\partial \mathbf{U}_g}{\partial \theta}, \quad (1)$$

where U and U_g are the horizontal and geostrophic winds, respectively (see Eliassen, 1949, 1962; Hoskins, 1975). With the relationship between horizontal acceleration and ageostrophic flow (U_{ag}) expressed by

$$\mathbf{U}_{ag} = \mathbf{U} - \mathbf{U}_g = f^{-1}\mathbf{k} \times \frac{d\mathbf{U}}{dt}, \quad (2)$$

combined with the assumption of adiabatic flow and neglect of $(\mathbf{U}\cdot\nabla_\theta \mathbf{k}) \times \mathbf{U}_g$, an alternative form for the ageostrophic velocity is

$$\mathbf{U}_{ag} = f^{-1}\Big[\underbrace{\mathbf{k} \times \frac{\partial \mathbf{U}_g}{\partial t_\theta}}_{\text{A}} + \underbrace{\mathbf{U}\cdot\nabla_\theta(\mathbf{k} \times \mathbf{U}_g)}_{\text{B}}\Big]. \quad (3)$$

Term A represents the isallobaric wind (Brunt and Douglas, 1928), while term B accounts for advective, inertial processes. A quasi-geostrophic form of (3) is

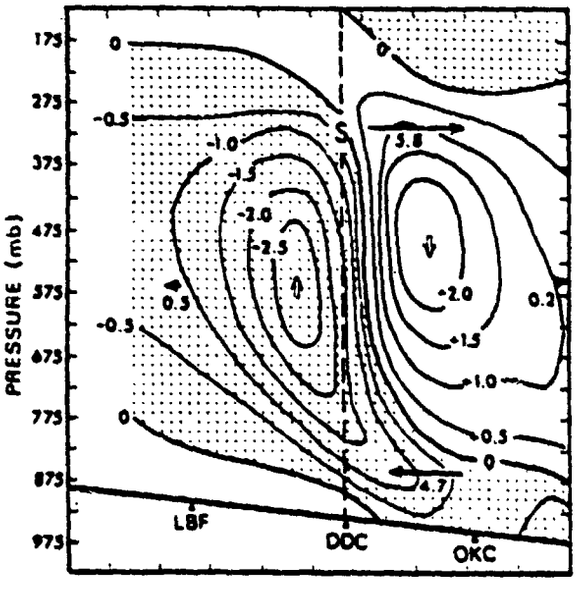


FIG. 2. Indirect circulation simulated by PE model mapped on cross section through exit region of a jet streak. Solid isopleths are vertical motion (μb s^{-1}) with upward motion stippled; arrows with centered magnitudes indicate horizontal transverse components (m s^{-1}) (from Cahir, 1971).

given by substituting $\mathbf{U}_\theta$ for the advecting velocity U. However, including the ageostrophic component in term B yields significant differences in the quantitative assessment of advective processes in the regions of strong gradients near fronts and jet streaks (Hoskins, 1975). Note also that with the use of the adiabatic constraint within the isentropic framework, the link between the ageostrophic motion and inertial processes is solely by the quasi-horizontal advection. If the expansion of the total derivative were made in Cartesian or isobaric coordinates, term B would include a mode for upper and lower tropospheric interaction through vertical advection.

Using isentropic coordinates, an explicit isallobaric mode (term A) may be isolated to identify the interaction between the upper and lower troposphere through mass and momentum adjustments. Substituting the hydrostatic geostrophic relation

$$\mathbf{U}_g = f^{-1}\mathbf{k} \times \nabla_\theta \psi \tag{4}$$

in (3), the isallobaric component of the ageostrophic wind is

$$\mathbf{U}_{ag_1} = -f^{-2}\nabla_\theta \frac{\partial \psi}{\partial t_\theta}, \tag{5}$$

where ψ is the Montgomery streamfunction. Integration of the hydrostatic equation from the earth's surface to an intermediate isentropic surface θ_L yields

$$\psi_L = \psi_s - \int_{\theta_s}^{\theta_L} c_p \left(\frac{p}{p_{00}}\right)^\kappa d\theta, \tag{6}$$

where p is pressure, $p_{00} = 1000$ mb, and κ is the ratio of the gas constant R for dry air to the specific heat of constant pressure c_p. By Leibniz's rule, the local derivative of (6) is

$$\frac{\partial \psi_L}{\partial t_\theta} = \frac{\partial \psi_s}{\partial t_\theta} + \frac{c_p \kappa}{p_{00}} \int_{\theta_s}^{\theta_L} \left(\frac{p}{p_{00}}\right)^{\kappa-1} \frac{\partial p}{\partial t_\theta} d\theta - c_p \left(\frac{p_s}{p_{00}}\right)^\kappa \frac{\partial \theta_s}{\partial t_\theta}. \tag{7}$$

With Poisson's equation and the definition of ψ, the time derivative of the surface value of the Montgomery streamfunction is

$$\frac{\partial \psi_s}{\partial t_\theta} = c_p \left(\frac{p_s}{p_{00}}\right)^\kappa \frac{\partial \theta_s}{\partial t_\theta} + \frac{c_p \theta_s \kappa}{p_{00}} \left(\frac{p_s}{p_{00}}\right)^{\kappa-1} \frac{\partial p_s}{\partial t_\theta}. \tag{8}$$

Substitution of (8) into (7) and a rearrangement of terms yields

$$\frac{\partial \psi_L}{\partial t_\theta} = R\left[\frac{T_s}{p_s}\frac{\partial p_s}{\partial t_\theta} + \int_{\theta_s}^{\theta_L} \left(\frac{p}{p_{00}}\right)^\kappa \frac{1}{p}\frac{\partial p}{\partial t_\theta} d\theta\right]. \tag{9}$$

With this result and (5), the isallobaric component of the ageostrophic wind ($\mathbf{U}_{ag_L}$) on θ_L is

$$\mathbf{U}_{ag_L}(\theta_L) = -\frac{R}{f^2}\left\{\nabla\left[\frac{T_s}{p_s}\frac{\partial p_s}{\partial t_\theta}\right] + \nabla\left[\int_{\theta_s}^{\theta_L}\left(\frac{p}{p_{00}}\right)^\kappa \frac{1}{p}\frac{\partial p}{\partial t_\theta} d\theta\right]\right\}. \tag{10}$$

The form of (10) reveals that the isallobaric wind in isentropic coordinates is determined by the gradients of the surface pressure tendency and the integrated pressure tendency between the earth's surface and θ_L. In a hydrostatic atmosphere, the tendency of the vertically integrated mass distribution determines the first term, while the second term is related to internal mass redistribution. Note that the isallobaric wind component may exist on an isentropic surface even if the surface pressure tendency is zero.

The relationship between the pressure tendencies and $\mathbf{U}_{ag_L}$ expressed in (10) couples tropospheric mass adjustment to the isallobaric wind on θ_L. The vertical integration of the mass continuity equation between arbitrary lower and upper isentropic surfaces (θ_2 and θ_1) yields

$$\frac{\partial p_2}{\partial t_\theta} = \frac{\partial p_1}{\partial t_\theta} + \int_{\theta_1}^{\theta_2}\left[\nabla_\theta \cdot \frac{\partial p}{\partial \theta}\mathbf{U} + \frac{\partial}{\partial \theta}\left(\frac{\partial p}{\partial \theta}\frac{d\theta}{dt}\right)\right] d\theta. \tag{11}$$

With the substitution of (11) into (10), the isallobaric ageostrophic component on θ_L can be linked to horizontal mass divergence and diabatic processes.

a. Mass adjustments associated with a propagating jet streak

The results from a hybrid isentropic and sigma coordinate numerical model are used to illustrate mass-momentum adjustments associated with the jet streak's propagation in a zonal channel. Following is a brief summary of the hybrid model described in detail by Uccellini *et al.* (1979). The hybrid model combines an isentropic representation of the free atmosphere with a sigma coordinate representation of the bottom 200 mb of the troposphere. The boundary conditions at the interface separating the model domains are matched exactly through use of the flux form of the governing equations for full interaction between the model domains without introducing spurious sources of mass, momentum and energy. A staggered vertical grid is used to maintain a smooth transition of the mass and wind fields without need for artificial adjustments. The longitudinal boundary conditions are cyclic while impervious, free-slip conditions are imposed on the latitudinal boundaries. The initial mass field is specified by a set of analytic functions which define the surface temperature gradient and the vertical lapse rates throughout the

entire model domain. The initial wind field is specified through the geostrophic relationship.

Areal average mass flux divergence was calculated for a $4\Delta x$ by $3\Delta y$ rectangle ($\Delta x = \Delta y = 2.75 \times 10^5$ m) in each of the four quadrants surrounding the jet streak 16 h into the model simulation. The averages were computed for the eight layers between 290 and 350 K with $\Delta\theta$ equal to 10 K.

With adiabatic conditions, the mass flux divergence is related solely to the horizontal mass flux divergence term in (11). The vertical distribution of the averaged total mass flux divergence (solid profile) illustrates a two-layer mass adjustment in the entrance and exit regions of the jet streak (Fig. 3). In the cyclonic exit and anticyclonic entrance quadrants, the upper tropospheric mass divergence is complemented by lower tropospheric mass convergence with the level of nondivergence located near 310 K. A reverse pattern exists in the anticyclonic exit and cyclonic entrance regions. The four-cell, two-layer pattern is consistent with Bjerknes' (1951) and Riehl *et al.* (1952) concepts of adjustments associated with a jet streak.

Some indication of the relative importance of geostrophic versus ageostrophic motion at this scale may be determined by partitioning the mass flux diver-

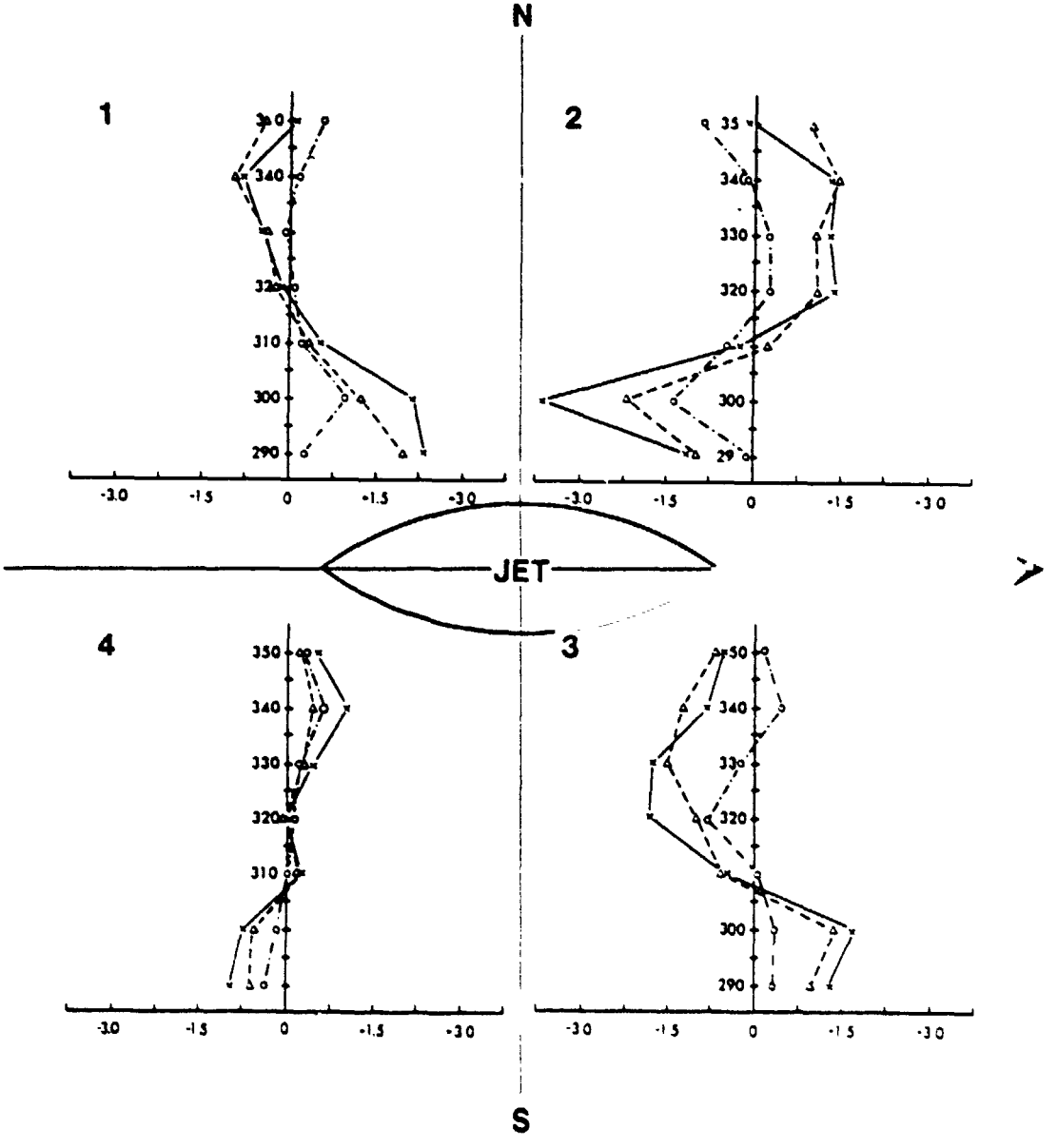


FIG. 3. Modeled vertical profiles for the averaged mass flux divergence ($\times 10^1$ g m^{-2} s^{-1}) [total (solid), ageostrophic (dashed), geostrophic (dot-dashed)] in four quadrants surrounding jet streak.

gence into a geostrophic and ageostrophic mode, where

$$\nabla_\theta\cdot(\rho J\mathbf{U}) = \nabla_\theta\cdot(\rho J\mathbf{U}_g) + \nabla_\theta\cdot(\rho J\mathbf{U}_{ag}), \quad (12)$$

where ρJ (density; Jacobian of transformation) is equal to $-(1/g)\partial p/\partial\theta$. The relative amplitude of the geostrophic (dot-dashed profile) versus the ageostrophic mode (dashed profile) shows that the ageostrophic mode of the mass flux divergence dominates at this scale of the mass adjustment (Fig. 3).

b. The upper tropospheric transverse component

An analysis of the components of the ageostrophic mass flux divergence (not shown) reveals that the cross-stream component $\partial(\rho J v_{ag})/\partial y$ generally dominates in the lower troposphere. In the upper troposphere, the cross-stream and alongstream ageostrophic components are of equal magnitude and of the same sign. Bjerknes (1951) emphasized the need for upper tropospheric ageostrophic components to provide for the cross-stream mass transport in the exit and entrance regions of the jet streak. In the exit region, the upper tropospheric, ageostrophic wind transports mass from the cyclonic toward the anticyclonic side of the jet. In the entrance region, the upper tropospheric ageostrophic wind transports mass from the anticyclonic toward the cyclonic side of the jet. At the level of the jet streak, where the horizontal variations of the pressure gradient force and wind are maximized, the advective inertial term in (3) tends to be larger than and offsets the isallobaric term in the forcing of a transverse ageostrophic component (Bjerknes, 1951). Fig. 4 illustrates the 340 K ageostrophic wind component defined by the inertial advective term. In the entrance region, the increase of the geostrophic wind along the direction of the streak yields a maximum ageostrophic transverse wind component of 4.5 m s^{-1} directed toward the cyclonic side. In the exit region, the decrease in the geostrophic wind along the axis of the jet yields a maximum ageostrophic transverse wind component of 8.5 m s^{-1} directed toward the anticyclonic side. This basic pattern propagates with the jet streak and remains quasi-steady.

c. The lower tropospheric return branch viewed as an isallobaric wind component

The opposite pattern of the areal average mass flux divergence below the 310 K level in each quadrant (Fig. 3) is an indication of the reversal of the lower tropospheric branches of the direct and indirect circulations from the upper branches illustrated in Fig. 4. In the lower troposphere, where advective velocities are small and inertia wind components are strongly damped by friction, the isallobaric wind modified by friction becomes a more important factor for low-level parcel accelerations and resultant ageostrophic flow (Hess, 1959; Young, 1973). In comparing the two contributing terms to U_{ag} in (3), the isallobaric wind (term A) is the dominant factor of the lower tropospheric transverse ageostrophic wind component for this model simulation. The largest, lower tropospheric isallobaric wind components are perpendicular to the axis of the jet streak and are opposite in direction to the inertial advective component of the upper branches (Fig. 5).

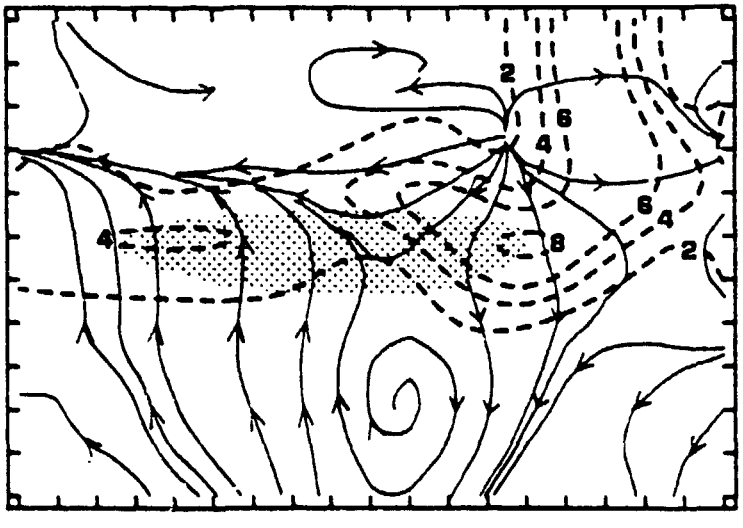


FIG. 4. Model simulated ageostrophic component related to the inertial-advective term [term B, Eq. (3)] on the 340 K surface. Streamline analysis (solid lines) is shown for the model domain up to the first interior row of grid points along the north and south boundaries; magnitudes (m s^{-1}) are shown as dashed lines. Shaded region represents jet streak (u component >40 m s^{-1}). The marks along the border represent 275 km grid spacing.

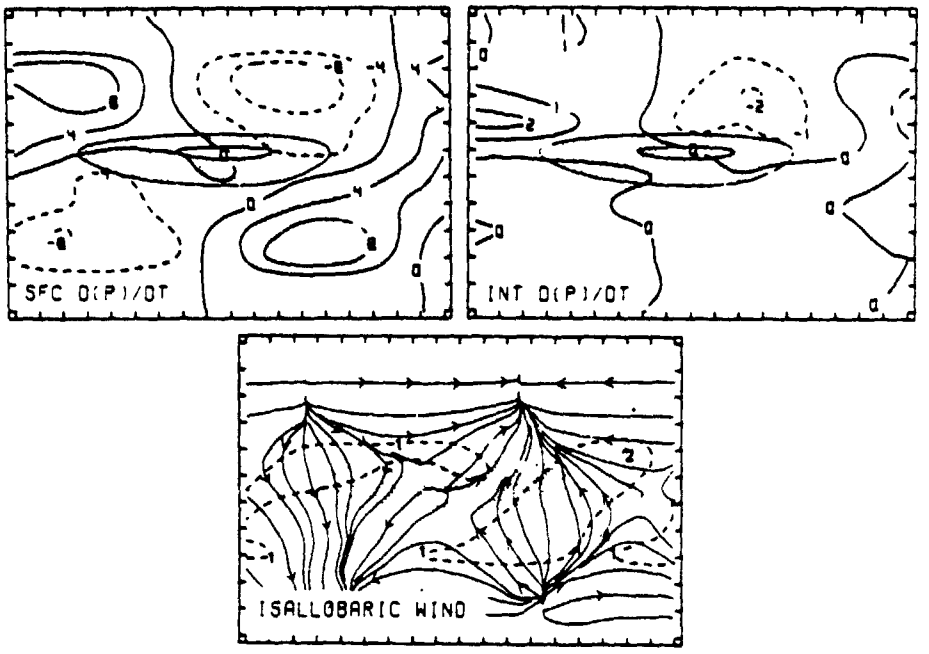


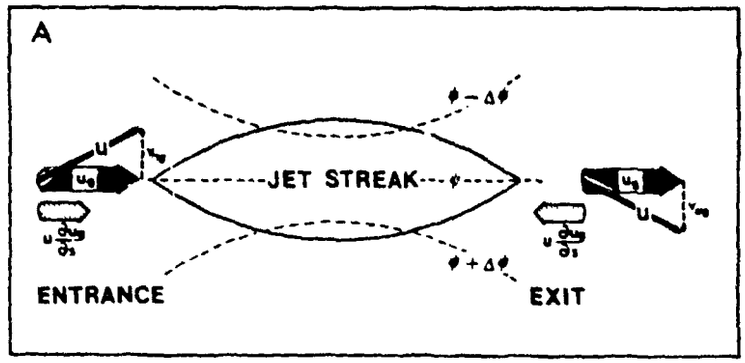
FIG. 5. (Above) The surface pressure tendency and integrated pressure tendency contributions to the ψ tendency from (9) evaluated on 300 K ($\times 10^{-3}$ m^2 s^{-2}). (Below) Streamline analysis is shown for isallobaric wind with magnitudes (m s^{-1}) as dashed line. See caption in Fig. 4 for additional details.

Maximum values of 1.5 m s^{-1} are located immediately ahead of and behind the streak. This basic pattern of the isallobaric wind components propagates with the jet streak and also remains quasi-steady at subsequent times.

Fig. 5 also illustrates the relation of the isallobaric wind to mass adjustments within the exit and entrance regions of the upper tropospheric jet streak. The surface and integrated pressure tendency terms in (9) were evaluated with θ_L equal to 300 K to determine the forcing of the isallobaric component within the return branches. Both terms display the basic four-cell pattern, with the surface pressure tendency term noticeably larger in these adiabatic experiments. In the cyclonic exit and anticyclonic entrance regions, pressure tendencies are negative while positive tendencies exist in the anticyclonic exit and cyclonic entrance regions. As specified by (10), the patterns of both pressure tendencies act to force the isallobaric wind directed toward the cyclonic and anticyclonic sides of the jet streak beneath the exit and entrance regions. These ageostrophic, isallobaric wind components represent lower tropospheric return branches of direct and indirect circulations.

d. Coupling the transverse branches to tropospheric mass adjustment

The principal components of the upper and lower tropospheric branches of the direct and indirect circulation are summarized in Fig. 6. In the entrance region of the upper tropospheric jet, confluent streamlines and downwind increase of the geostrophic wind are linked with the inertial advective process and are associated with a transverse ageostrophic component directed toward the cyclonic side of the jet streak (Fig. 6A). In the exit region, diffluent streamlines and a downwind decrease of the geostrophic wind are associated with an ageostrophic component directed towards the anticyclonic side. The upper tropospheric, cross-stream mass transports that are a result of the transverse components act to force the lower tropospheric isallobaric winds which represent the lower branches of the transverse circulations. Beneath the entrance region of the upper tropospheric jet streak, the upper tropospheric mass transport reduces the slope of the lower tropospheric isentropic surface (as p increases to the left and decreases to the right), decreases the ψ gradient, and yields an ageostrophic component on θ_L directed to the right of the current (Fig. 6B). Beneath the exit region, the upper tropospheric mass transport increases the slope of the isentropic surface and the ψ gradient to yield an ageostrophic component on θ_L directed to the left of the current (Fig. 6C). These ageostrophic components in the upper and lower troposphere represent the *transverse components* of the direct and indirect circulations, while the total wind component (U in Fig. 6) represents the *total* upper and lower tropospheric branches of the circulations.



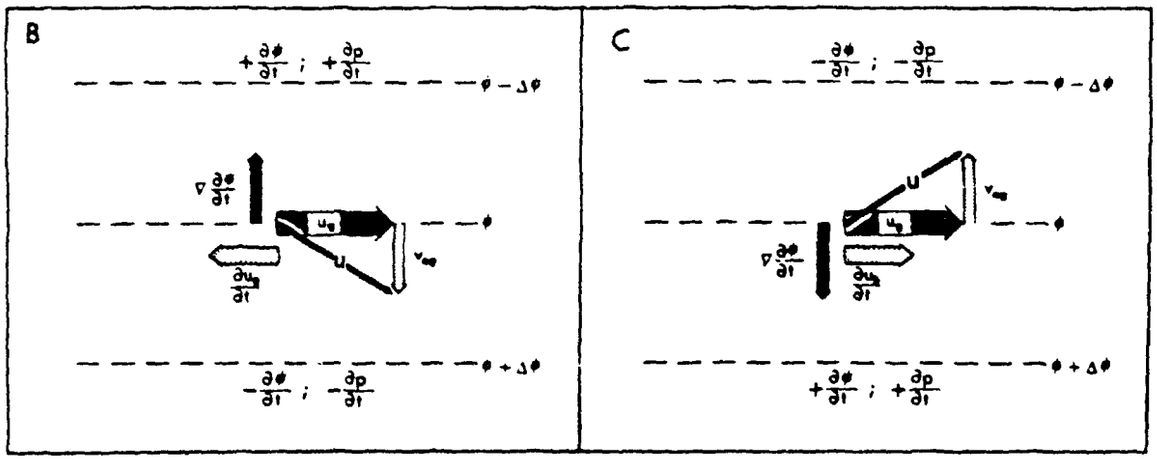


FIG. 6. (A) Schematic of transverse component (v_{ag}) forced by along-stream variations of pressure gradient force at jet streak level. (B), (C) Isallobaric wind (v_{ag}) on lower tropospheric isentropic surface (θ_L) beneath the entrance (B) and exit (C) regions of the jet streak. Geostrophic wind is V_g, total wind is V, pressure is p and the alongstream coordinate is s.

The schematic in Fig. 7 depicts a cross-sectional view of the lower tropospheric isallobaric wind forced by adiabatic mass adjustments in the exit region of a jet streak. As a jet streak (J) propagates toward the vertical cross section in Fig. 7A, the upper tropospheric ageostrophic component transports mass toward the anticyclonic side of the jet. The mass flux convergence on the anticyclonic side and divergence on the cyclonic side above θ_L increases the pressure on and beneath θ_L to the right of the jet, and decreases the pressure to the left of the jet. By (9) the mass adjustment within the upper branches above θ_L increases the magnitude of the pressure gradient force on and beneath θ_L and by (10) concurrently forces development of an isallobaric wind that represents a return branch of the indirect circulation (Fig. 7B). With the concurrent development of the return branch, mass is transported toward the cyclonic side of the jet below θ_L. This response represents a mutual mass-momentum adjustment and is an integral part of the sinking and lifting of isentropic surfaces beneath θ_L to higher and lower pressures on the anticyclonic and cyclonic sides of the jet, respectively. The two-layer adjustment depicted in Fig. 7 stems from Rossby's adjustment concepts for a stratified fluid (Rossby, 1938, 1949; University of Chicago, 1947).

A schematic of the influence of cross-stream variation of diabatic mass transport in increasing the magnitude of the pressure gradient on θ_L is depicted in Figs. 7C and 7D. An increase in the isallobaric wind on θ_L will occur with either diabatic heating to the right or diabatic cooling to the left. The heating vertically transports mass from below to above θ_L, while cooling transports mass from above θ_L to below θ_L.

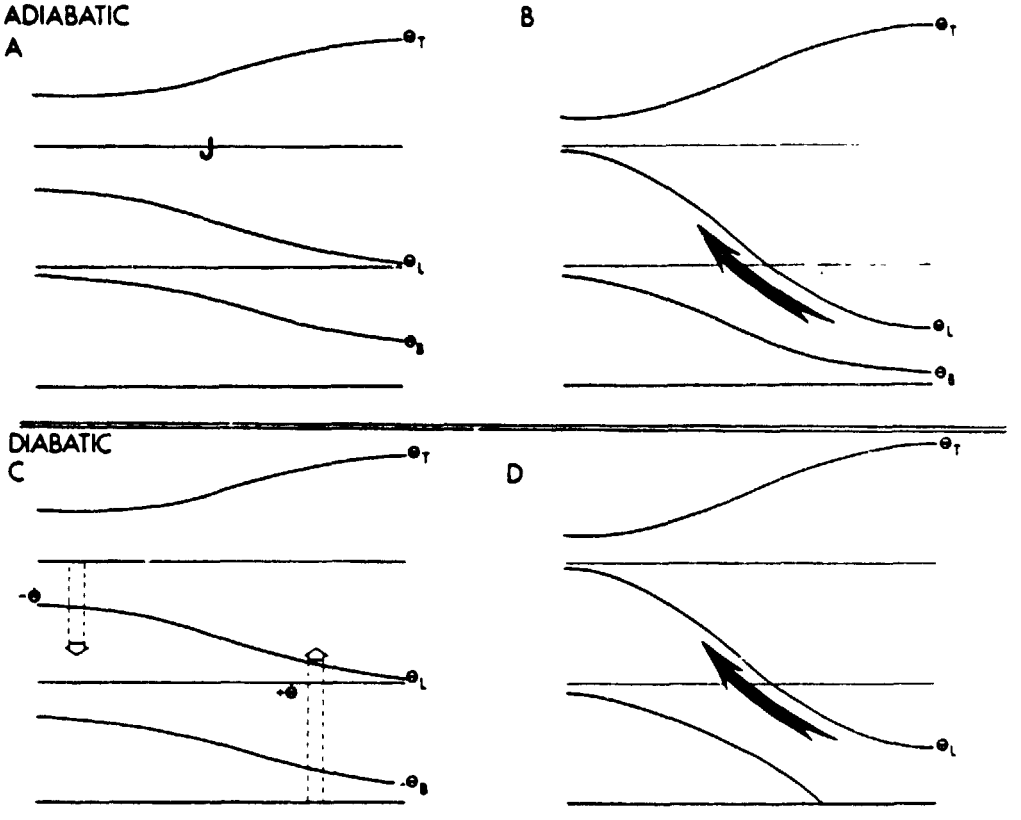


FIG. 7. (A), (B) Vertical cross section normal to axis of jet (J) showing adiabatic mass transports in exit region of approaching jet streak and resultant isallobaric wind (heavy arrow in B). (C), (D) Diabatic mass transports contributing to isallobaric wind (heavy arrow in D)

(Fig. 7C). Since diabatic mass transports are vertical, the surface pressure tendencies are not directly affected by the first term in (9) which reduces to zero. However, with the assumed diabatic mass transports depicted in Fig. 7C the vertically integrated pressure tendency term in (9) becomes positive to the right and negative to the left for isentropic surfaces below θ_L. As a consequence, the intensification of the pressure gradient force generates a diabatic component of the isallobaric wind along θ_L, directed toward the left (Fig. 7D). It is likely that the release of latent energy *may at times* force such a component if the cumulus convection develops in warm moist regions to the right of the jet streak. The diabatic mass transport related to sensible heat flux could also be important as θ_L approaches the planetary boundary layer to the right of the jet. If the diabatic heating is maximized to the left of the jet, as is implied by Fig. 1, the magnitude of the isallobaric component would be reduced. Since the preliminary model experiments were for a dry atmosphere, these effects could not be evaluated to determine their relative importance.

4. Case study: 10–11 May 1973

Evidence of coupling between upper and lower tropospheric jet streaks by mutual mass-momentum adjustments and the subsequent generation of convective instability is now presented from a case study. The severe weather outbreak over Indiana and Ohio on 10–11 May 1973 was selected for study since the lack of cyclogenesis in this case allows for the isolation of mass adjustments and the response of the lower tropospheric wind forced primarily by a jet streak.

The synoptic conditions preceding the development of the severe convective storms are described in this section. The propagation of an upper tropospheric jet streak and two-layered mass adjustment are studied to establish the forcing of the lower tropospheric isallobaric wind on the 300 K surface. A diagnostic trajectory analysis is then applied to link the isallobaric wind and the development of a low-level jet on the 300 K surface. Finally, the development of the LLJ and its interaction with the upper tropospheric jet streak are related to the development of the convective storm system in eastern Indiana and Ohio.

a. Surface and radar analyses

On the morning (1200 GMT) of 10 May 1973, a diffuse low-pressure system was located in the Great Lakes and Hudson Bay regions with a weak cold

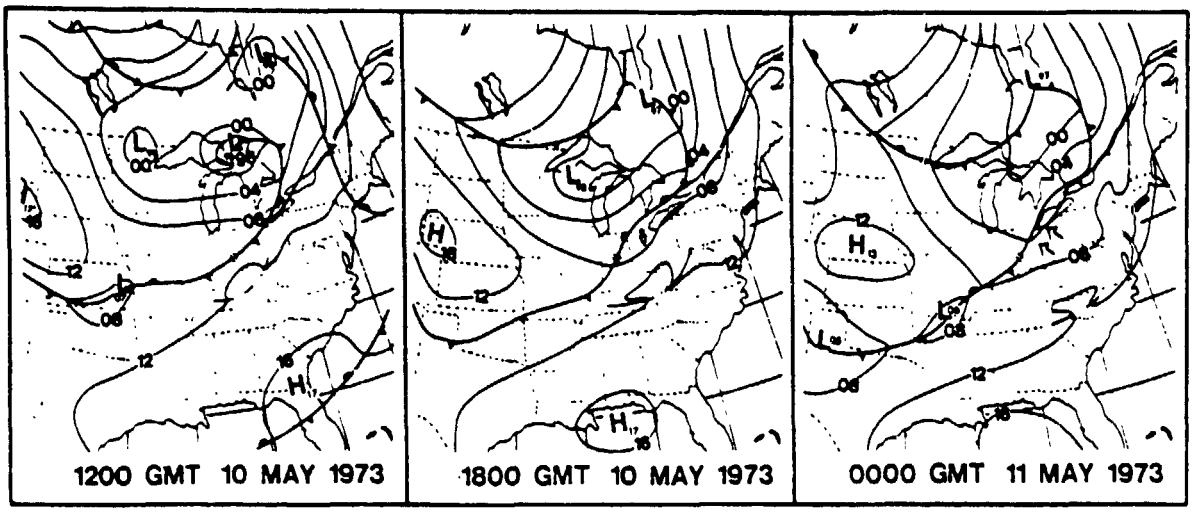


FIG. 8. Surface isobar analyses (mb, 08 = 1008 mb) for 1200 GMT 10 May through 0000 GMT 11 May 1973.

front extending from east of Lake Huron southwestward through southern Illinois and Missouri where a stationary front connected to a weak low in Kansas (Fig. 8). At 1800 GMT 10 May, the low in the Great Lakes region remained diffuse but the characteristics of the frontal zone changed. The portion of the front from north of Lake Erie to central Indiana moved northward and was analyzed as a warm front. The frontal zone from Illinois to Kansas propagated slowly southward. By 0000 GMT 11 May the frontal zone in Ohio ceased moving northward and assumed the characteristics of a cold front accelerating to the east, with severe weather located ahead of the front in Ohio.

The convective storms began developing in eastern Indiana and northwest Ohio ahead of the front by 1740 GMT and rapidly evolved into several squall lines which propagated eastward during the afternoon (Fig. 9). A separate squall line developed in southern Illinois—southeast Missouri and propagated to the Kentucky—Tennessee border by 2340 GMT in association with the front extending into Missouri. The storms in Indiana and Ohio were responsible for 9 confirmed tornados and 15 funnel cloud reports, while the southern system produced 5 confirmed tornados and 6 funnel clouds.

b. Upper air analyses

The upper air analyses in this study are subjective. The isentropic and pressure surface analyses are cross checked with 15 vertical cross sections to incorporate the detailed vertical resolution of individual rawinsonde ascents into the wind, pressure and moisture analyses on isentropic surfaces and wind, temperature and moisture analyses on pressure surfaces (see Shapiro, 1970; Fig. 2). The cross-check technique also insures consistency between analyses on pressure and isentropic surfaces for each time period.

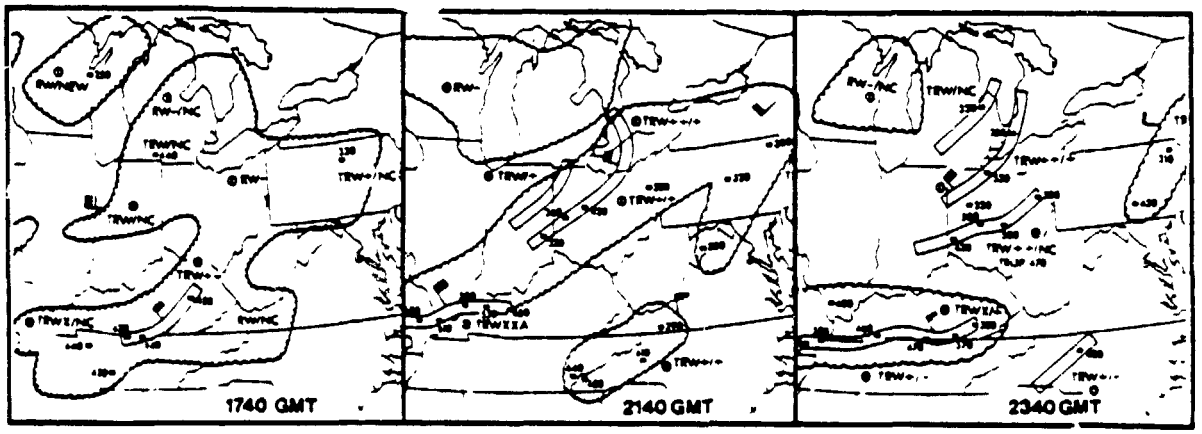


FIG. 9. Radar depictions, 10 May 1973. Cloud tops in hundreds of feet (430 = 43000 ft), RW rainshowers, TRW thundershowers.

At 1200 GMT 10 May, three jet streaks on the 330 K surface (Fig. 10A) were embedded within a general westerly flow stretching from the West Coast to the Great Lakes region. Of primary interest is the upper tropospheric jet streak with a maximum wind speed slightly over 60 m s^{-1} extending from Nebraska to Iowa. The isotach analysis on the 300 K surface (Fig. 10B) illustrates general westerly flow with no evidence of a LLJ perpendicular to the axis of the upper tropospheric jets at this time. The 850 mb isotach analysis for 1200 GMT 10 May (not shown) did indicate that relatively weak, south to southwest winds on the order of 5 m s^{-1} existed over west central Illinois. An extensive band of moisture with mix

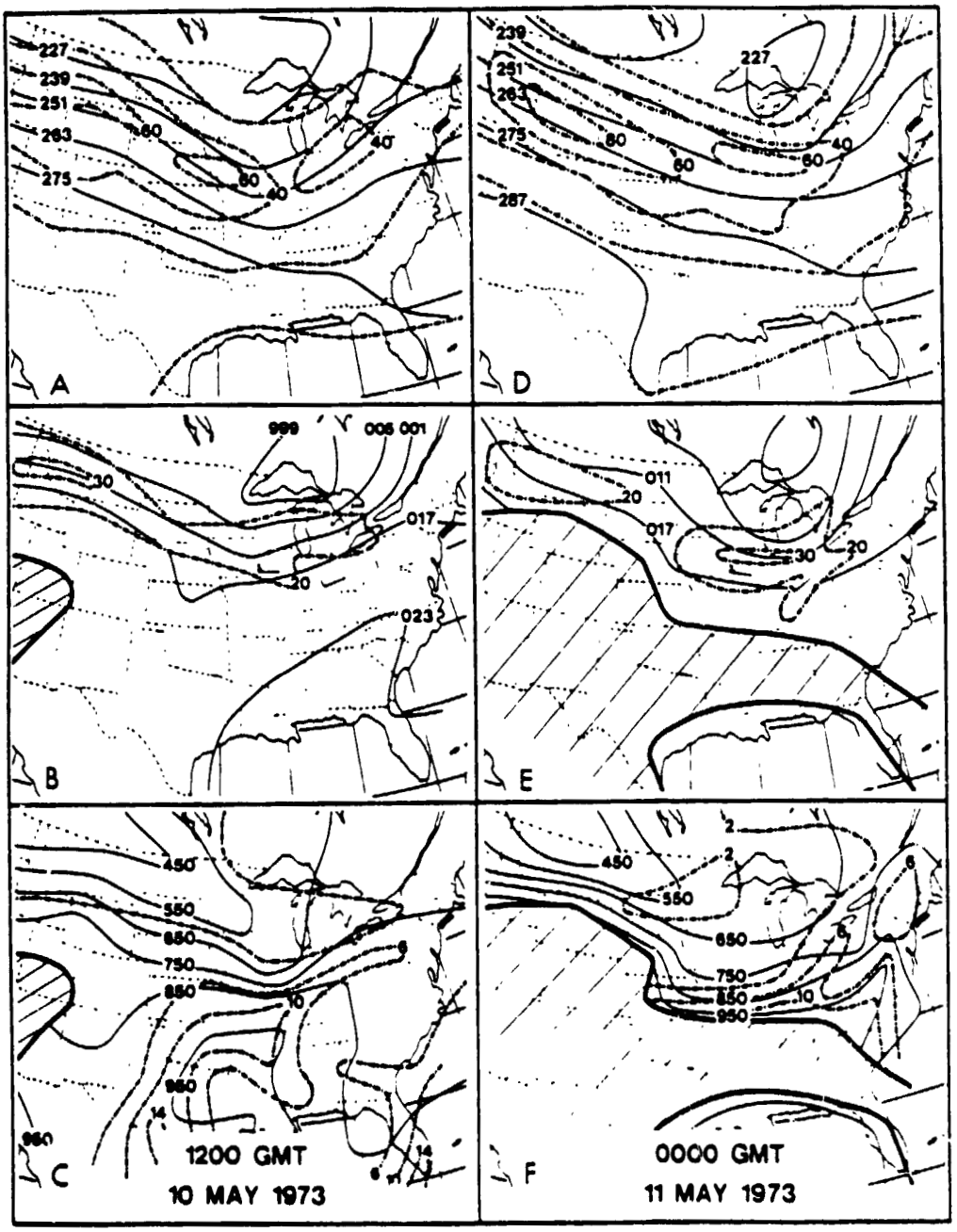


FIG. 10. Isentropic analyses for 1200 GMT 10 May and 0000 GMT 11 May 1973. (A), (D) 330 K isotach and ψ; (B), (E) 300 K isotach and ψ, with isotachs (dashed, m s^{-1}) and ψ (solid, 275 = 3 275 × 10^6 m^2 s^{-2}); (C), (F) 300 K pressure (solid, mb) and moisture (dashed, g kg^{-1}).

ing ratios exceeding 6 g kg^{-1} extended northward from the Gulf of Mexico into the Central Plains and then eastward into southwestern Pennsylvania (Fig. 10C). The moisture tongue from southern Illinois to Pennsylvania was parallel to the isobars (a measure of the slope of the 300 K surface) and generally remained below the 800 mb level.

By 0000 GMT 11 May, the jet streak over the Central Plains intensified slightly and propagated eastward, extending from eastern Iowa to northwest Ohio (Fig. 10D). At the same time a low-level jet with southwesterly winds >20 m s^{-1} existed on the 300 K surface and extended from the 950 mb level in northern Kentucky to 700 mb in western Pennsylvania (Fig. 10E). The moisture tongue sharpened considerably by 0000 GMT, stretched from Kentucky to Ohio coincident with the axis of the LLJ, and was directed up the isentropic surface to the 700 mb level in northeast Ohio (Fig. 10F). Another wind maximum >30 m s^{-1}, as supported by information gained by cross-section analysis, is also evident on the 300 K surface in northern Illinois (Fig. 10E) and apparently represents the downward extension of the upper tropospheric jet streak analyzed on the 330 K surface. This wind maximum was characterized by a generally westerly flow nearly parallel to the ψ contours and to the 650 mb isobar. The westerly wind maximum was therefore significantly different from the ageostrophic LLJ which accelerated toward the lower ψ values and ascended the 300 K isentropic surface over Ohio.

By 1200 GMT 11 May, the upper tropospheric jet streak, which propagated eastward to Pennsylvania, was orientated more from the southwest to the northeast, had weakened to 45 m s^{-1} and was generally not well defined. On the 300 K surface, both the lower tropospheric wind maximum and moisture tongue broadened considerably, lessened their intensity, and shifted east and south from the 0000 GMT positions.

c. The lower tropospheric wind maximum at 0000 GMT 11 May

The development of the LLJ extending from Kentucky to Ohio at 0000 GMT 11 May is the most important feature of this case study. Its relative position on the 850 mb surface (Fig. 11B) within the exit region of the upper tropospheric jet streak at the 300 mb level (Fig. 11A) is nearly identical to Newton's schematic (Fig. 1). While the analysis on the 850 mb pressure surface clearly illustrates the LLJ, the 300 K isentropic analysis (Fig. 10E) better depicts the vertical extent of the LLJ and suggests a direct interaction with the upper tropospheric jet streak analyzed on the 330 K surface. The LLJ was directed up the 300 K isentropic surface, extending from the 950 mb level in Kentucky to 750 mb in Pennsylvania. The LLJ axis was at a noticeable angle to the ψ contours over Ohio and at nearly a 30° angle to the height contours on the 850 mb surface (Fig. 11B), indicating that the wind maximum was significantly ageostrophic and accelerating. With the lower tropospheric wind maximum directed up the sloped isentrophic surfaces and toward the cyclonic side of the upper tropospheric jet streak, it appears that the LLJ was embedded within the return branch of the indirect circulation. It must be emphasized that this lower tropospheric jet streak has different characteristics from the low-level jets common to the southern Great Plains during the spring and summer months (Section 2). The LLJ in this case was maximized in the late afternoon and extended up toward the 700 mb level, unlike the nocturnal jets which at-

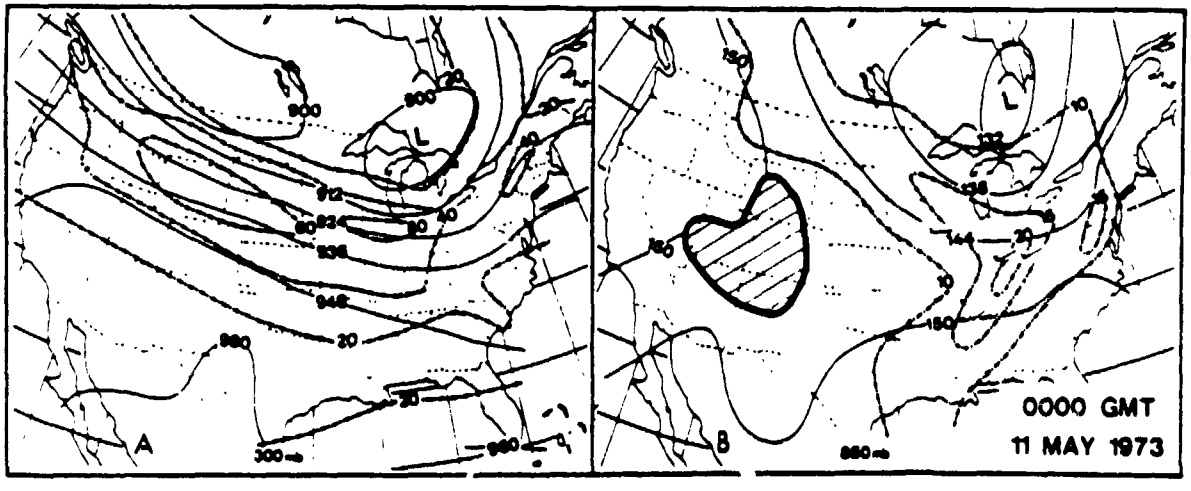


FIG. 11. Pressure surface analyses for 0000 GMT 11 May 1973. (A) 300 mb isotach (dashed, m s^{-1}) and geopotential height (solid, 924 = 9240 m). (B) 850 mb isotachs and geopotential height (solid, 144 = 1440 m).

tain maximum intensity in the early morning and are usually restricted to the planetary boundary layer. The remainder of this case study is concerned with isolating the two-layer mass adjustment associated with the forcing of the return branch and studying the processes responsible for the formation of the LLJ.

d. Analysis of the two-layer mass adjustment

The mass adjustments associated with the jet streak propagating into Ohio at 0000 GMT 11 May are isolated by calculating the tendencies of the hydrostatic pressure difference between isentropic surfaces (Δp) using an overlapping polynomial technique developed by Whittaker and Petersen (1977) and discussed in the Appendix. Fig. 12 illustrates the Δp tendencies within two layers, 340 to 300 K and 300 K to the earth's surface, during the 12 h period (1200 GMT 10 May–0000 GMT 11 May). The pattern of the tendencies for the upper and lower layers is very similar to fields predicted for the jet streak by the hybrid model (Section 3). Within the exit region, upper tropospheric mass convergence and divergence occur on the anticyclonic and cyclonic sides

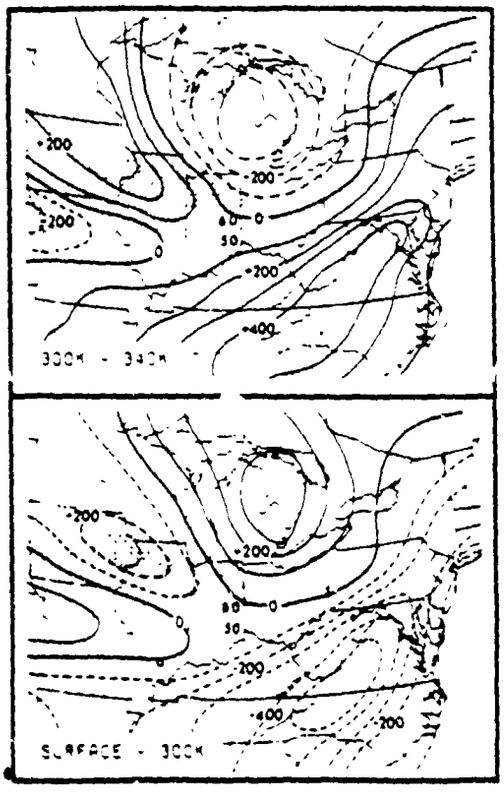


FIG. 12. Mass tendencies ($\times 10^{-1}$ g m^{-2} s^{-1}) for (A) 340 to 300 K layer and (B) from earth's surface to 300 K for period between 1200 GMT 10 May and 0000 GMT 11 May. Light shaded region represents jet streak on 330 K isotachs greater than 50 m s^{-1}; dark shaded region, isotachs greater than 60 m s^{-1} at 0000 GMT 11 May 1972.

yielding the positive and negative tendencies, respectively. The pattern is reversed in the lower layer. Although the mass transports were not explicitly calculated for the diagnostic case study, the two-layer pattern revealed in Fig. 12 is apparently a result of the mass transports associated with the transverse circulations in the exit and entrance regions as simulated by the hybrid model (Section 3).

Supporting evidence of the two-layer mass adjustment ahead of the propagating upper tropospheric jet is illustrated in the vertical cross sections from Sault Ste. Marie, Michigan (SSM) through Ohio to Athens, Georgia (AHN) (Fig. 13). At 1200 GMT 10 May, the mass distribution measured by Δp between the 305 and 325 K isentropic surfaces is relatively uniform along the entire cross section. The jet core located at the 250 mb level between Flint, Michigan (FNT), and Dayton, Ohio (DAY), is identified with the northeastward propagating jet streak centered over southeast Michigan (Fig. 10A). By 0000 GMT 11 May, the jet streak originally located over Nebraska propagated eastward and entered the plane of the vertical cross section near the 375 mb level between FNT and DAY. The mass distribution displays significant changes during this 12 h period with Δp between 305 and 325 K decreasing 200 mb to the left and increasing 150 mb to the right of the streak. The Δp changes in this case replicate the schematic of the two-layer mass adjustment associated with the cross-contour components of the indirect transverse circulations in the exit region (Fig. 7A and 7B). However, the change in Δp in the lower layer to the right of the jet could also be related to diabatic fluxes associated with planetary boundary layer processes and latent heat release (Figs. 7C and 7D).

e. The lower tropospheric isallobaric wind

The impact of the two-layer mass adjustment on the isallobaric wind is determined through an evaluation of (9) and (10) for the 300 K surface. The analysis of the integrated pressure tendency term in (9) is presented in Fig. 14. A four-cell pattern is apparent with the integrated pressure tendency increasing to the right and decreasing to the left of the jet streak in the exit region. In the entrance region a weaker reversed cross-stream variation of the integrated pressure tendency is present. The cross-stream variation of the integrated pressure tendencies is forced by the two-layer mass adjustment. In the exit region, mass convergence above 300 K and mass divergence below forced the 300 K surface to higher pressure on the anticyclonic side. On the cyclonic side, the reversed pattern forced the 300 K surface to lower pressure. In the entrance region, the opposite pattern of the integrated pressure tendencies indicates a reversed two-layer adjustment.

The influence of the integrated and surface pressure tendency terms on the 300 K isallobaric wind as

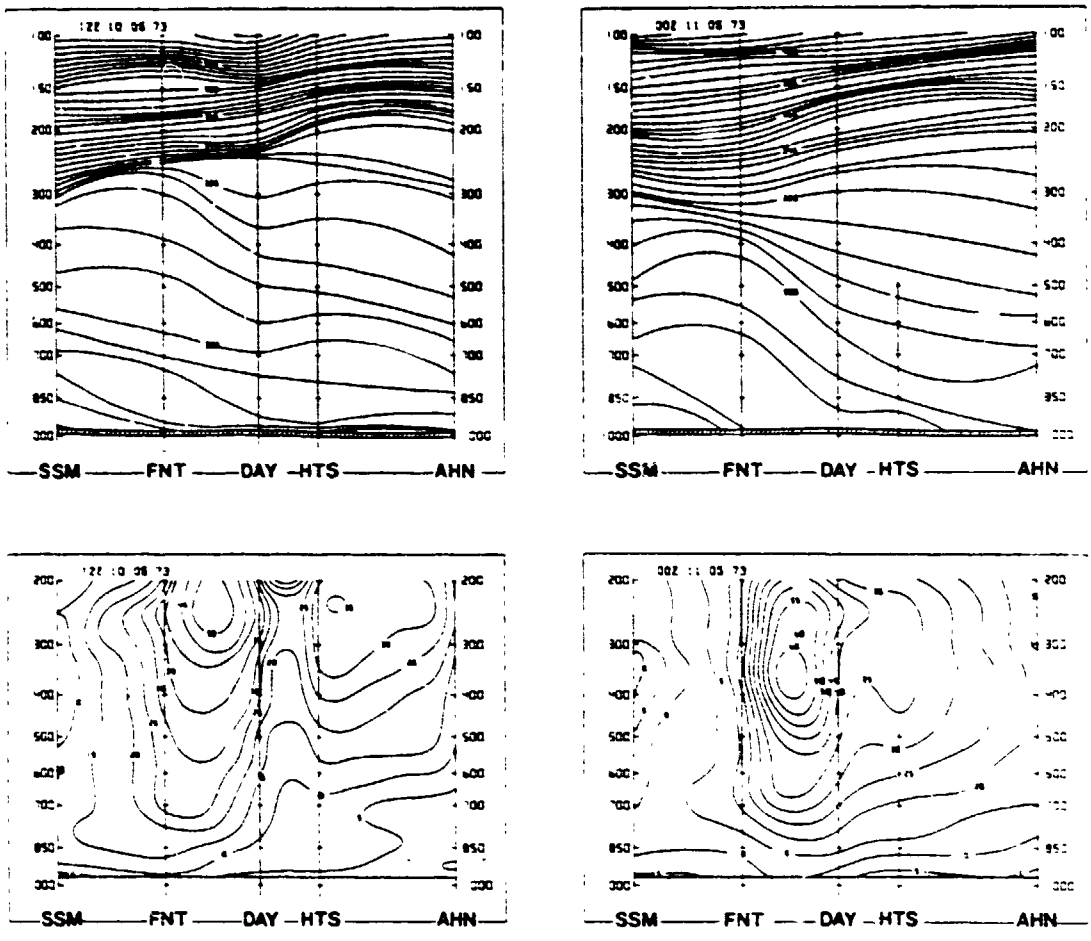


FIG. 13. Vertical cross sections (1200 GMT 10 May 1973 left, 0000 GMT 11 May 1973 right). Potential temperatures (K) top and isotachs (m s^{-1}) bottom.

specified by (10) is presented in Fig. 15. The isallobaric wind vectors (Fig. 15A) converged toward the cyclonic wave in Michigan and toward the mean frontal position through western Ohio and Kentucky (Fig. 8). The maximum isallobaric wind vectors from southern Wisconsin to northeast Indiana are above the 650 mb level and reflect the deepening mid-tropospheric wave in the Great Lakes region. The isallobaric wind beneath the exit region of the upper tropospheric jet streak was orthogonal to the axis of the streak and directed toward the cyclonic side. This ageostrophic component represents the lower branch of the indirect circulation that was evident in the numerical simulation (Section 3).

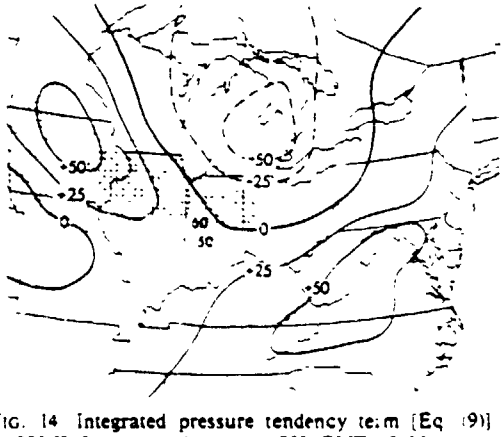


FIG. 14. Integrated pressure tendency term [Eq. (9)] with θ_L = 300 K for period between 1200 GMT 10 May and 0000 GMT 11 May ($\times 10^{-4}$ m^2 s^{-3}). See caption for Fig. 12 for additional details.

The southerly isallobaric wind within the lower branch of the indirect circulation in north central Ohio is of particular interest since it coincides with the position of the LLJ over Ohio at 0000 GMT 11 May. The isallobaric wind component in this region was dominated by the integrated pressure tendency term (Fig. 15B) over the surface pressure tendency term (Fig. 15C). The positive integrated pressure tendency to the right of the jet in the exit region and negative tendency to the left (Fig. 14) was forced by the internal two-layer mass adjustment within the exit region.

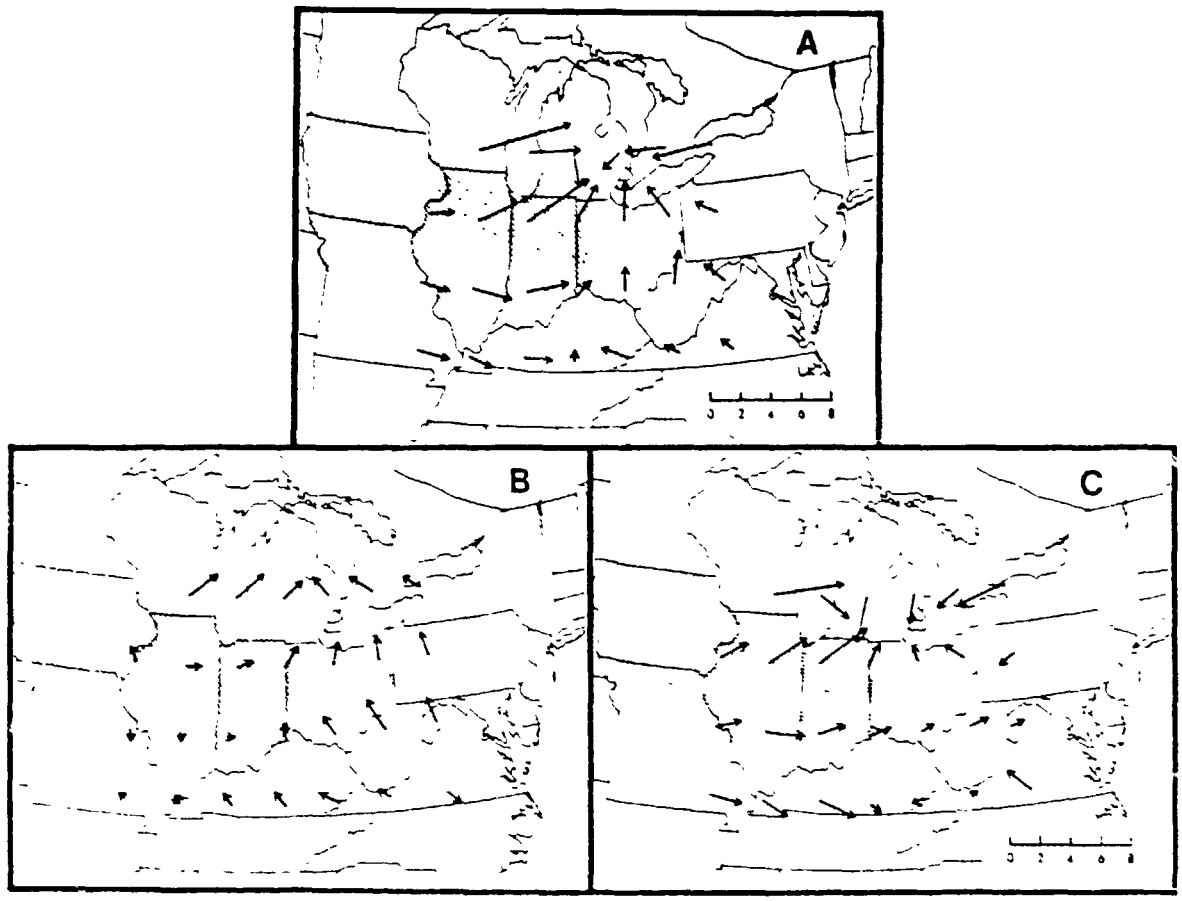


FIG. 15. Evaluation of isallobaric wind [Eq. (10)] and contributing terms for 12 h period 1200 GMT 10 May to 0000 GMT 11 May 1973. (A) vector representation for negative tendency of pressure gradient force [multiplication by f^{-1} yields isallobaric wind], (B) negative integrated pressure tendency term, (C) negative surface pressure tendency term. Magnitude is scaled to lower right of each figure ($\times 10^{-4}$ m s^{-2}). Shaded region represents jet streak at 0000 GMT 11 May, see Fig. 12 for details.

f. A diagnostic trajectory analysis

Trajectories were computed and geostrophic and ageostrophic components of the wind were analyzed to determine the impact of the isallobaric wind on the development of the LLJ. The trajectories are calculated by a diagnostic technique that is based on Greenspan's (1972, 1973) discrete model formulation and which utilizes the primitive form of the inviscid equations of motion (Petersen and Uccellini, 1979). The success of this trajectory technique, like others, is susceptible to analysis errors and violation of the assumptions. However, it should be emphasized that the discrete model approach nearly conserves energy (Petersen and Uccellini, 1979) and that the trajectories represent an accurate response to the calculated $\dot{\sigma}$ tendencies insofar as the inviscid and adiabatic assumptions are satisfied. The final parcel velocities were compared to the subjective isotach analysis for 0000 GMT 11 May to check the accuracy of the trajectory computations and were generally within 3 m s^{-1} and 20° of the subjective isotach and isogon analyses from 0000 GMT 11 May.

Six trajectories on the 300 K surface that were selected from over 40 trajectories calculated by the discrete model approach are illustrated in Fig. 16. Trajectories A and D depict the parcel movement through the lower extension of the upper tropospheric jet streak located at approximately the 650 mb level. The acceleration of parcel D to 30 m s^{-1} over northern Indiana agrees with the observed winds at 0000 GMT 11 May (Fig. 10E). Trajectories C, E and F illustrate the parcels that originate in the lower troposphere and accelerate northeastward through Ohio and toward Pennsylvania. The parcels entering southern Ohio turn toward the north and accelerate up the isentropic surface (toward lower pressure) in the region where the southerly isallobaric wind component was the largest (Fig. 15A) and where the LLJ existed at 0000 GMT 11 May (Fig. 10E). Parcels C and E, for example, accelerated to wind speeds of 23 and 18 m s^{-1} between 1800 and 0000

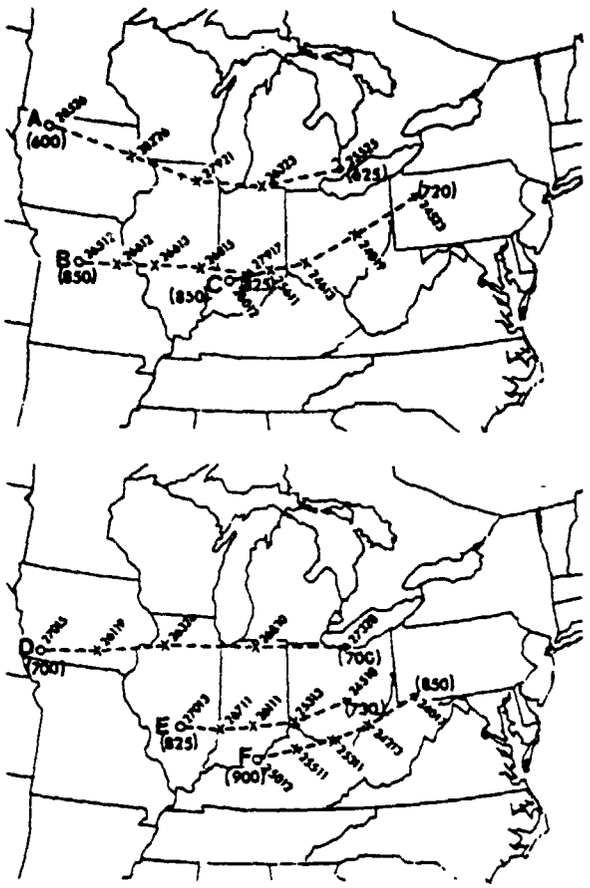


FIG. 16. Parcel trajectories calculated on 300 K isentropic surface for the 12 h period 1200 GMT 10 May to 0000 GMT 11 May 1973; wind velocity is indicated at 3 h positions (24017 is 240° at 17 m s^{-1}); initial and final pressures (mb) are indicated in parentheses.

GMT, respectively, and ascended from below 825 mb to above the 750 mb level. The upward vertical motion indicated by these parcel trajectories reflects the increasing slope of the 300 K surface and the southerly wind component directed toward lower pressure on the 300 K surface (Figs. 10C, 10E and 10F).

The total, geostrophic and ageostrophic wind components and the isallobaric ageostrophic components for parcels C, E and D are presented in Table 1. The geostrophic and isallobaric winds were determined by specifying grid-point ψ values and ψ tendencies for each time by the overlapping polynomial technique (Appendix), calculating geostrophic and isallobaric wind components at the grid points and interpolating to the parcel position. The ageostrophic wind is calculated from the difference between the total and geostrophic winds.

Parcels C and E were chosen since their trajectories pass through the region in which the LLJ formed by 0000 GMT 11 May. The total u and v wind components for C increased 4.6 and 4.4 m s^{-1}, respectively, as it moved into central Ohio at 2100 GMT. Parcel E underwent a similar acceleration as it entered central Ohio at 0000 GMT 11 May. The increases in the v component for C and E are largely due to the increases in v_{ag}. For parcel C, v_{ag} increased to 4.1 m s^{-1} in central Ohio by 2100 GMT and for E to 4.6 m s^{-1} by 0000 GMT. The intensification of v_{ag} is to a large degree accounted for by the increase of the isallobaric v component to 3.0 m s^{-1}. The increase in v_{ag} as parcels C and E entered Ohio is responsible for the computed acceleration of the u component through inertial rotation. The increases in u_g as parcels C and E entered Ohio is consistent with the strengthening north–south ψ gradient on the 300 K surface in that region.

TABL. 1. Total, geostrophic, ageostrophic and isallobaric wind components for parcel trajectories C, E, D in Fig. 17.

Hour	Position	Parcel velocity	u	v	u_g	v_g	u_{ag}	v_{ag}	u_{ag_3}	v_{ag_3}
				Parcel C						
1200 GMT	SW Indiana	260 12	11.8	2.0	10.0	3.2	1.8	−1.2	0.5	0.0
1800 GMT	SW Ohio	246 13	11.9	5.2	13.0	4.0	−1.1	1.2	0.0	2.0
2100 GMT	E central Ohio	240 19	16.5	9.6	16.5	5.5	0.0	4.1	−0.3	3.0
0000 GMT	NW Pennsylvania	245 23	20.8	9.7	16.5	8.0	4.3	1 7	−0.5	1.3
				Parcel E						
1200 GMT	S Illinois	270 13	13 0	0.0	14.5	3.1	−1.5	−3.1	0 9	−0 5
1800 GMT	S central Indiana	261 11	10.8	1.7	10.5	1.2	0 3	0.5	1.0	0.0
2100 GMT	SW Ohio	253 13	12.4	3.8	14.0	1.5	−1.6	2.3	1.0	2.0
0000 GMT	S central Ohio	245 18	16.3	7.6	18.0	3.0	−1.7	4.6	0.7	3 0
				Parcel D						
1200 GMT	SW Iowa	270 15	15.0	0.0	—	—	—	—	—	—
1800 GMT	N Illinois	263 26	25.8	3.1	22.0	0.0	3.8	3 1	3.0	1.5
2100 GMT	N Indiana	268 30	29.9	1.0	25.5	−1.0	4 4	2.0	2.0	2.0
0000 GMT	N Ohio	273 28	27.9	−1.5	24 0	5 0	3.9	−6.5	−0.7	1 7

Parcel D illustrates the response of super-geostrophic parcels entering northern Ohio upon exiting the lower extension of the upper tropospheric jet streak. As D entered northern Ohio (0000 GMT), it turned slightly to the south with the v_{ag} decreasing from 2.0 to -6.5 m s^{-1}. The negative v_{ag} depicts the response of D after movement through the jet streak in northern Indiana into a region of a weaker pressure gradient force and indicates that for supergeostrophic parcels exiting the streak, the inertial effect [term B in Eq. (3)] dominates the positive isallobaric contribution in northern Indiana and Ohio. Comparing supergeostrophic parcel D with the initially subgeostrophic parcels C and E shows that the isallobaric component dominates the ageostrophic wind in the lower troposphere to the south and southwest, while the inertial component dominates the middle troposphere to the west and northwest. The patterns of motion and acceleration for these trajectories from the northwest and southwest infer deformation and/or convergence in the vicinity of the front near which the severe weather developed.

The analysis of the trajectories indicates that the increases in the u and v components for lower tropospheric parcels entering the region of the LLJ over southwest Ohio are consistent with the expected response of the lower tropospheric wind to the mass adjustments in the exit region of an upper tropospheric jet streak and the increasing pressure gradient force on lower tropospheric isentropic surfaces. The isallobaric component associated with the increased pressure gradient force accounts for over half of the increase in the southerly ageostrophic wind in Ohio. The southerly ageostrophic wind forces a corresponding increase in the u component through inertial rotation. The result of the increases in the v_{ag} and u components is the formation of the LLJ directed toward the northeast, oriented at a significant angle to the axis of the upper tropospheric jet streak (Fig. 11).

g. Implications of coupled jet streaks for convective storm development

The impact of the coupled jets on creating conditions favorable for the deep convection in this case study was determined by computing the meridional moisture and sensible heat transports along a cross section from southern New York westward to Nebraska (Fig. 17). With the large increase of the lower tropospheric meridional component associated with the development of the LLJ, the maximum northward moisture transport over Ohio increased from 30 to 102 g m^{-1} s^{-1}, or an equivalent increase in the transport of latent energy from 7.6 to 25.7 $\times$ 10^{4}J m^{-1} s^{-1} and lowered to the layer within which the LLJ formed (Fig. 17B). For the same 12 h period, the maximum northward sensible heat transport over Ohio increased from 12.6 to 28.1 $\times$ 10^{5}J m^{-1} s^{-1} in conjunction with the development of the LLJ.

The impact of the increased southerly wind in the lower troposphere and at the earth's surface in the region where the LLJ developed is also illustrated by the surface equivalent potential temperature (θ_e) analyses (Fig. 18). The θ_e analyses are based on hourly surface data from nearly 100 Midwest reporting stations. The increased temporal and spatial resolution lend valuable supporting evidence to the interpretation that the LLJ increased the lower tropospheric moisture and heat transports into the inception area of the severe convective storms. A θ_e tongue rapidly intensified and, by 1800 GMT 10 May, extended northeastward into Ohio coinciding with the axis of the LLJ at 0000 GMT 11 May. Although the daytime insolation can significantly increase the surface θ_e, the increase by insolation would tend to be uniform at any given latitude. The shape and preferred location of the narrow θ_e tongue coinciding with the axis of the LLJ points to the importance of the accelerated advective process in the rapid evolution of the θ_e field in southeastern Indiana and western Ohio. As in other cases (Means, 1952, 1954; Bonner, 1966), the narrow region of accelerated moisture and temperature advections in the boundary layer is due to the increased lower tropospheric winds associated with the development of the LLJ.

While the development of the LLJ increased the lower tropospheric θ_e in Ohio by rapid northward transport of sensible heat and moisture, the lower portion of the upper tropospheric jet streak that extended down to the 600–700 mb layer (see 300 K analysis, Fig. 10E) transported cooler and drier air into the middle troposphere. The effect of the differential transports was to reduce θ_e by about 4 K within the 500 to 750 mb layer for the Dayton, Ohio soundings from 1200 GMT 10 May to 0000 GMT 11 May, while lower tropospheric θ_e values increased by 4 to 12 K. By 0000 GMT 11 May, the Dayton sounding indicated convective instability from the earth's surface to 500 mb, and that the level of free convection was lowered from 770 to 850 mb, while the equilibrium level at the top of the positive energy area of latent instability increased from 680 to 270 mb. The net effect of the differential sensible heat and moisture transports associated with the coupled jet streaks was to create conditions favorable for the formation of deep convective storms that developed within the exit region of the upper tropospheric jet streak by 0000 GMT.

5. Summary

The interaction between upper and lower tropospheric jet streaks is an important factor in the development of organized severe convective storm systems. The intersection of jet axes in the exit re-

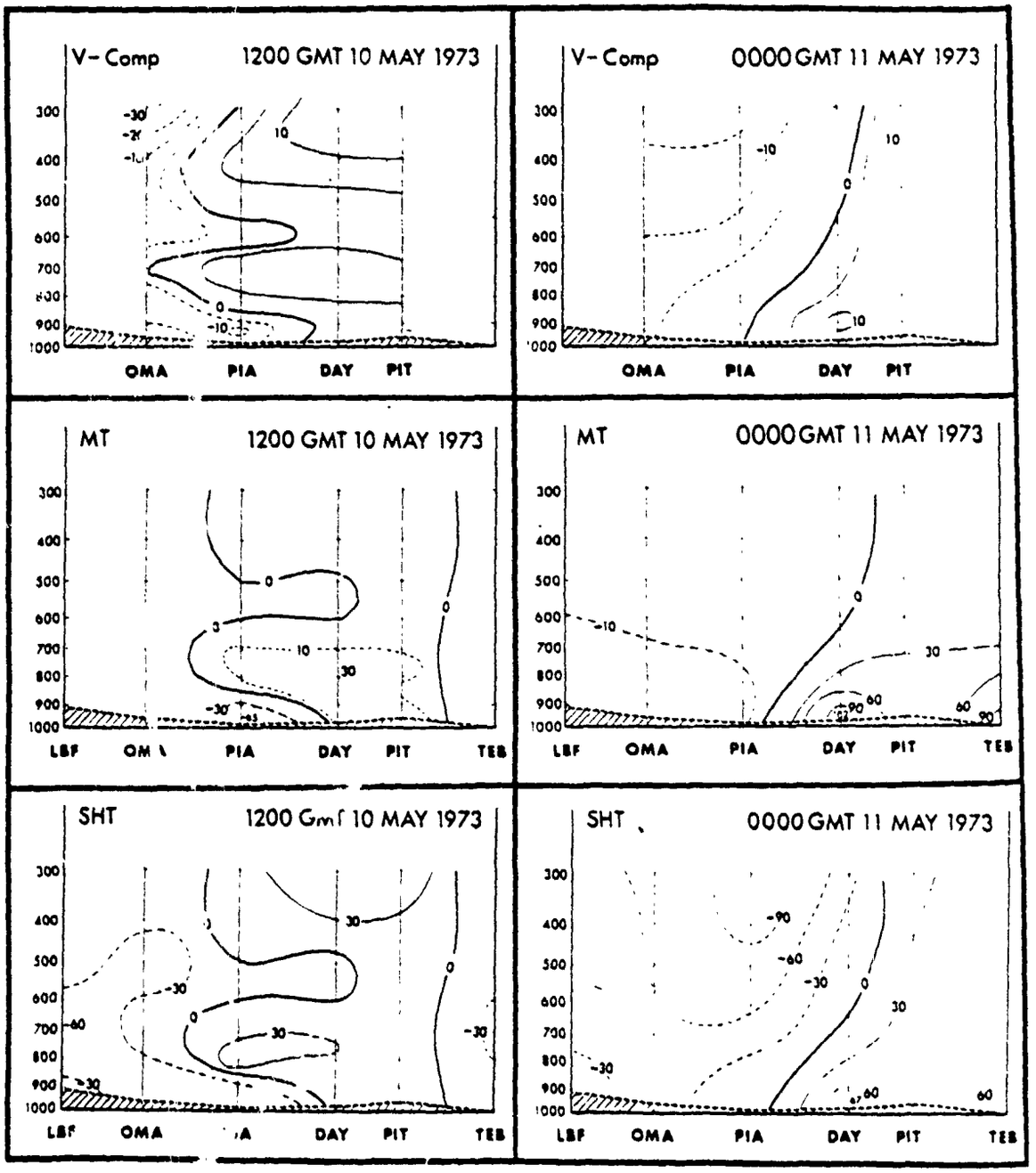


FIG. 17. Vertical cross sections of (A) meridional wind component (m s^{-1}) from Omaha, Nebraska (OMA) to Pittsburgh, Pennsylvania (PIT), (B) meridional moisture transport (MT, g m^{-1} s^{-1}) and sensible heat transport (SHT, $\times 4.2 \times 10^4$ J m^{-1} s^{-1}) from North Platte, Nebraska (LBF) to Teterboro, New Jersey (TEB). Northward directed fluxes are positive (solid), southerly fluxes are negative (dashed).

gion of the upper tropospheric jet and the veering of the wind with height consecutively destabilizes the atmosphere through differential moisture and temperature advections (Miller, 1955; Petterssen, 1956; Newton, 1967). In most studies, the tendency has been to treat each jet streak as a separate entity, although Reiter (1969) suggested that in some cases upper and lower tropospheric jet streaks are not separate entities. This paper has presented a basis for the dynamical processes responsible for coupling upper and lower tropospheric jet streaks through mutual mass-momentum adjustments and transverse

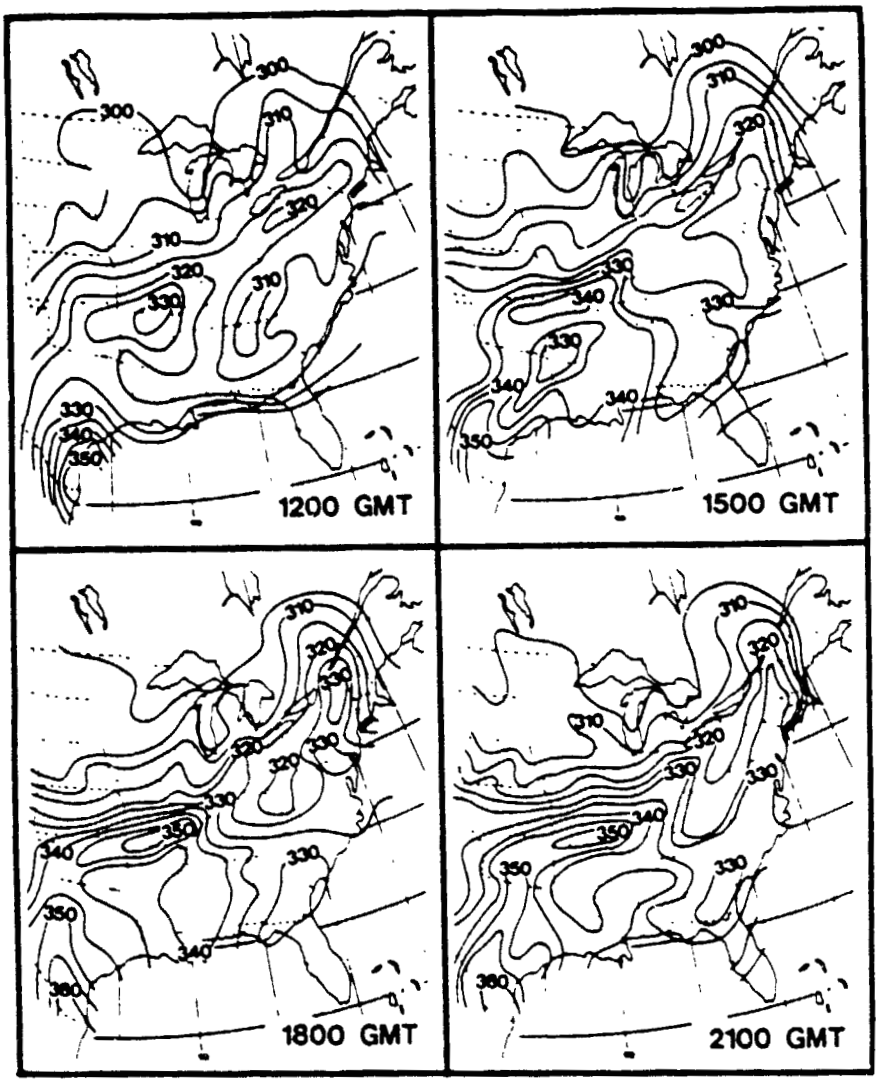


FIG. 18. Surface equivalent potential temperature (θ_e) analyses for 10 May 1973, based on FAA and Weather Service hourly surface data.

circulations within the exit region of an upper level jet streak. The forcing of the transverse circulations was investigated by the combined application of the geostrophic momentum approximation and isallobaric equation in isentropic coordinates

A hybrid numerical model is used to simulate the mass adjustment associated with a jet streak propagating in a zonal channel and the response of the lower troposphere. The model simulation depicts a basic two-layer mass adjustment in the exit and entrance regions of the jet streak that was related primarily to the ageostrophic mode. At the level of the jet streak, the inertial advective component of the geostrophic momentum approximation to the ageostrophic wind dominated the upper branches of the transverse circulations. In the lower troposphere, the isallobaric component of the ageostrophic wind dominated the lower branch of the transverse circulations. The numerical results verified that the lower tropospheric isallobaric winds, representing the transverse components of the return branches of the direct and indirect circulations, are forced by the two-layer mass adjustment accompanying the propagation of the upper level jet streak.

In the case study, a similar two-layer mass adjustment was isolated in the exit region of the upper tropospheric jet streak and linked to the increased pressure gradient force and the resultant isallobaric component within the return branch of an indirect circulation. The mass adjustment and isallobaric wind were then related to the formation of a low-level jet. The LLJ, located beneath the exit region

of the upper tropospheric jet streak, was embedded within the lower branch of an indirect circulation. The LLJ was directed up sloped isentropic surfaces toward the cyclonic side of the upper level jet and was noticeably ageostrophic. A diagnostic analysis of atmospheric trajectories verifies that a southerly isallobaric wind was an important component of the LLJ in this case. The southerly isallobaric wind was largely responsible for the observed intensification of the meridional ageostrophic wind component and forced a corresponding acceleration of the u component through inertial rotation, while u_g intensified through the increased pressure gradient force. The increases in v_{ag} and u resulted in the formation of the LLJ directed toward the northeast at a significant angle to the axis of the upper tropospheric jet streak. Since the mass adjustments are responsible for the isallobaric ageostrophic wind, the upper tropospheric jet streak and LLJ were coupled by the two-layer mass adjustment associated with the jet streak's propagation. These results, which emphasize momentum generation in the lower troposphere as a result of a three-dimensional mass-momentum adjustment process, are in contrast to Ninomiya's (1971) contention that convection-induced vertical transport of momentum is responsible for the development of a LLJ within the exit region of an upper level jet streak. A downward transport of the upper level jet streak's momentum cannot account for the orthogonal momentum component of the LLJ that yields the difference in wind direction between upper and lower tropospheric jet streaks.

The case study also isolated how the coupling of the upper and lower tropospheric jet streaks yields conditions favorable for the development of severe convective storms within the *exit region* of the *upper tropospheric jet*. The isallobaric wind component of the LLJ that is orthogonal to the axis of the upper tropospheric jet streak was the primary reason for the LLJ being at a significant angle to the jet's axis. The dominance of the isallobaric component in the lower troposphere combined with the increased importance of the inertial advective effects in the middle and upper troposphere resulted in veering of the winds with height and differential moisture and sensible heat transports. The LLJ rapidly transported moisture and sensible heat within the lower troposphere northward into the inception area, while the lower extension of the upper tropospheric jet streak transported cooler and drier air eastward within the middle troposphere. The net effect of the differential transports was to generate convective instability from the earth's surface to the 500 mb level, to lower the level of free convection and to raise the equilibrium level capping latent instability, all of which are conducive to the formation of deep convective storms. The concept that the development of conditions favorable for deep convection occurs through mutual mass-momentum adjustments associated with a propagating jet streak provides an underlying basis for the criteria used in severe weather forecasts (Fawbush *et al.*, 1951; Fawbush and Miller, 1953, 1954).

The primary emphasis of this research has been to determine the role of propagating jet streaks and mass-momentum adjustments in creating conditions favorable for the development of severe convective storm systems. However, several important questions remain for future research. The influence that latent and sensible heating, and subsequent diabatic mass transports, have on the coupling of the lower tropospheric wind response to the propagating upper tropospheric jet streak should be examined through numerical experiments and additional case studies. Another important problem that should be considered in future research is an apparent non-steady aspect of the coupling between upper and lower tropospheric jet streaks. In this case, the sounding data did not indicate a LLJ within the exit region of the upper tropospheric jet streak at 1200 GMT 10 May. Also, the surface θ_e analysis for 1200 GMT did not reveal any distinct θ_e tongues in Illinois and Iowa. The development of a distinct θ_e tongue into southern Indiana and western Ohio during the day (Fig. 18), the reversal of the movement of the frontal zone in that area during the same period (Fig. 8), and the existence of the LLJ at 0000 GMT 11 May suggest that the LLJ did not move with the propagating upper tropospheric jet streak. Rather, the LLJ seemed to develop within the exit region between 1200 GMT 10 May and 0000 GMT 11 May as the upper tropospheric streak approached the surface front extending through Ohio and Indiana. The development of the LLJ appears to represent an acceleration of the transverse indirect circulation with respect to the propagating jet streak. Eliassen (1951) emphasized that the intensity of meridional circulations will increase with reduced stability. In this case, as the upper tropospheric jet streak approached the warmer air east of the front, the jet encountered a region in which the mid-tropospheric stability was significantly reduced on the cyclonic and anticyclonic sides of the streak (indicated by large Δp between 325 and 305 K isentropes in the 1200 GMT vertical cross section in Fig. 13). Perhaps the reduced stability in the region into which a jet streak propagates is a key factor for the intensification of the indirect circulation and a subsequent increase in the magnitude of the low-level winds. The examination of these questions could provide additional insight into the scale interactive processes that establish convective instability and initiate and maintain the severe convective storm systems.

Acknowledgments. We express our appreciation to Drs. J. A. Young, W. Schwerdtfeger and L. H.

Horn for their useful comments during the course of this research. We thank the reviewers for their valuable comments which helped to clarify portions of the manuscript. We also thank Thomas M. Whittaker for assisting in the computer programming, Jim Vergin and Jack Katzfey for assisting in data analysis and John Stremikis, Jean Johnson and Nancy Malz for preparing the manuscript. This research was sponsored by the Atmospheric Science Section, National Science Foundation, under Grants ATM75-23223 and ATM77-22976.

APPENDIX

The Calculation of the Isallobaric Wind and Mass Tendencies for the Case Study

The ψ and pressure tendencies in (11) and (12) are evaluated utilizing the "overlapping polynomial technique" developed by Bleck and Haagenson (1968) and expanded on by Whittaker and Petersen (1977). The method was designed for spatial objective analyses but can also be used to generate temporal series. In an example from Whittaker and Petersen (see Whittaker and Petersen, 1977; Fig. 3), two second-order polynomials are fitted between four data points. In the region between the second and third points, the two curves are merged using linear distance weighting to yield the resultant third-order polynomial. The method maintains a smooth transition between adjacent polynomials and insures that first derivatives can be calculated along the entire curve. In our application, the four data points represent sounding data from 0000 GMT 10 May through 1200 GMT 11 May 1973 at 12 h intervals. The period of interest in this case study is between the second and third points, 1200 GMT 10 May and 0000 GMT 11 May, respectively. This techhique is utilized to generate time series which are then used 1) to calculate the pressure at the midpoint of the series (1800 GMT 10 May) at each grid point for estimating the two contributing terms to U_{ag_i} in (10); 2) to specify the ψ tendencies in the trajectory analysis; and 3) to calculate the isallobaric component for parcel trajectories in Table 1 by computing grid-point tendencies of pressure gradient force at 3 h intervals with Δt set equal to 2 h, which are then interpolated to the parcel positions.

REFERENCES

Beebe, R. G., and F. C. Bates, 1955: A mechanism for assisting in the release of convective instability. *Mon. Wea. Rev.*, **83**, 1–10.

Bjerknes, J., 1951: Extratropical cyclones. *Compendium of Meteorology*, T. F Malone: Ed. Amer. Meteor. Soc., 577–598.

Blackadar, A. K., 1957: Boundary layer wind maxima and their significance for the growth of nocturnal inversions. *Bull. Amer. Meteor. Soc.*, **38**, 283–290.

Bleck, R., and P. L. Haagenson, 1968: Objective analysis on isentropic surfaces. NCAR Tech. Note, NCAR-TN-39, 27 pp.

Bonner, W. D., 1966: Case study of thunderstorm activity in relation to the low-level jet. *Mon. Wea. Rev.*, **94**, 167–178.

——, 1968: Climatology of the low level jet. *Mon. Wea. Rev.*, **96**, 833–850.

Browning, K. A., and T. W. Harrold, 1970: Air motion and precipitation growth at a cold front. *Quart. J. Roy. Meteor. Soc.*, **96**, 369–389.

Brunt, D., and C. K. M. Douglas, 1928: *Memo. Roy. Meteor. Soc.*, 3, No. 22.

Cahir, J. J., 1971: Implications of circulations in the vicinity of jet streaks at subsynoptic scales. Ph.D. thesis, Pennsylvania State University, 170 pp.

Cahn, A., 1945: An investigation of the free oscillations of a simple current system. *J. Meteor.*, **2**, 113–119.

Danielsen, E. F., 1974: The relationship between severe weather, major dust storms and rapid cyclogenesis. *Synoptic Extratropical Weather Systems*. M. Shapiro, Ed., National Center for Atmospheric Research, 215–241.

Eliassen, A., 1949: The quasi-static equations of motion with pressure as independent variable. *Geofys. Publ.*, **17**, 5–44.

——, 1951: Slow thermally or frictionally controlled meridional circulation in a circular vortex. *Astrophy. Norv.*, **5**, 19–60.

——, 1962: On the vertical circulation in frontal zones. *Geofys. Publ.*, **24**, 147–160.

Fawbush, E. J., and R. C. Miller, 1953: The tornado situation of 17 March 1951 *Bull. Amer. Meteor Soc.*, **34**, 139–145.

——, and ——, 1954: The types of air masses in which North American tornadoes form. *Bull. Amer Meteor Soc.*, **35**, 154–165.

——, —— and L. G. Starrett, 1951. An empirical method of forecasting tornado development. *Bull. Amer. Meteor Soc.*, **32**, 1–9

Gerhardt, J R., 1962. An example of a nocturnal low-level jet stream. *J Atmos. Sci.*, **19**, 116–118.

——, 1963: Mesoscale association of a low-level jet stream with a squall-like cold-front situation. *J. Atmos. Sci.*, **20**, 49–55

Greenspan, D., 1972. A new explicit discrete mechanics with applications. *J Franklin Inst.*, **294**, 231–240.

——, 1973: *Discrete Models*. Addison-Wesley, 105–107.

Hess, S. L., 1959: *Introduction to Theoretical Meteorology*. Holt, Rinehart and Winston, 225–227.

Hoecker, W. H., 1963: Three southerly low-level jet streams delineated by the Weather Bureau special pibal network of 1961. *Mon. Wea. Rev.*, **91**, 573–582.

Hoskins, B. J., 1975: The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmos. Sci.*, **32**, 233–242.

Izumi, Y., 1964: The evolution of temperature and velocity profiles during breakdown of a nocturnal inversion and a low-level jet. *J Appl. Meteor.*, **3**, 70–82.

——, and M. L. Barad, 1963. Wind and temperature variations during development of a low-level jet. *J. Appl. Meteor.*, **2**, 668–673.

Johnson, D. R., 1970: The available potential energy of storms. *J. Atmos. Sci.*, **27**, 727–741.

Lettau, H. H., 1967: Small to large scale features of boundary structures over mountain slopes. *Proc. Symp. Mountain Meteor.*, Colorado State University, 1–74.

Ludlam, F H., 1963: *Severe Local Storms. A Review. Meteor. Monogr.*, No. 27. Amer. Meteor. Soc., 1–30.

Means, L. L., 1952: On thunderstorm forecasting in the central United States. *Mon. Wea. Rev.*, **80**, 165–189

——, 1954: A study of the mean southerly wind maxima in low levels associated with a period of summer precipitation in the middle west. *Bull. Amer. Meteor. Soc.*, **35**, 166–170

Miller, J. E., 1955. Intensification of precipitation by differential advection. *J Meteor.*, **12**, 472–477

Murray, R., and S. M. Daniels, 1953. Transverse flow at entrance and exit to jet streams. *Quart. J Roy. Meteor. Soc.*, **79**, 236–241.

Namias, J., and P. F. Clapp, 1949: Confluence theory of the high tropospheric jet stream. *J. Meteor.*, **6**, 330–336.

Newton, C. W., 1954: Frontogenesis and frontolysis as a three-dimensional process. *J. Meteor.*, **11**, 449–461.

——, 1959: Axial velocity streaks in the jet stream: ageostrophic "inertial" oscillations. *J. Meteor.*, **16**, 638–645.

——, 1963: Dynamics of severe convective storms. *Meteor. Monogr.*, No. 27, Amer. Meteor. Soc., 33–55.

——, 1967: Severe convective storms. *Advances in Geophysics*, Vol. 12, Academic Press, 257–303.

Ninomiya, K., 1971: Mesoscale modification of synoptic situations from thunderstorm development as revealed by ATS III and aerological data. *J. Appl. Meteor.*, **10**, 1103–1121.

Paegle, J., and G. E. Rasch, 1973: Three-dimensional characteristics of diurnally varying boundary-layer flows. *Mon. Wea. Rev.*, **101**, 746–756.

Palmén, E., and C. W. Newton, 1969: *Atmospheric Circulation Systems*. Academic Press (see Chaps. 4, 5, 8, 9 and 13).

Petersen, R. A., and L. W. Uccellini, 1979: The computation of isentropic atmospheric trajectories using a "discrete model" approach. *Mon. Wea. Rev.*, **107**, 566–574.

Petterssen, S., 1956: *Weather Analysis and Forecasting*, Vol. 2. McGraw-Hill, 191–195.

Reiter, E. R., 1963: *Jet-Stream Meteorology*. The University of Chicago Press (see Chaps. 4 and 6).

——, 1969: Tropopause circulation and jet streams. *World Survey of Climatology*, Vol. 4., *Climate of the Free Atmosphere*. D. F. Rex, Ed., 85–193.

Riehl, H., et al., 1952. *Forecasting in the Middle Latitudes*. *Meteor. Monogr.*, No. 5, Amer. Meteor. Soc., 80 pp.

Rossby, C. G., 1938: On the mutual adjustment of pressure and velocity distributions in certain simple current systems II. *J. Mar. Res.*, **7**, 239–263.

——, 1949: On the nature of the general circulation of the lower atmosphere. *The Atmosphere of the Earth and Planets*. G. P. Kuiper, Ed., The University of Chicago Press, 16–48.

Sawyer, J. S., 1956: The vertical circulation at meteorological fronts and its relation to frontogenesis. *Proc. Roy. Soc. London.*, **234A**, 346–362.

Shapiro, M. A., 1970: On the applicability of the geostrophic approximation to upper-level frontal-scale motions. *J. Atmos. Sci.*, **27**, 408–420.

——, 1975: Simulation of upper-level frontogenesis with a 20-level isentropic coordinate primitive equation model. *Mon. Wea. Rev.*, **103**, 591–604.

Staley, D. O., 1960: Evaluation of potential vorticity changes near the tropopause and the related vertical motions, vertical advection of vorticity and transfer of radioactive debris from stratosphere to troposphere. *J. Meteor.*, **17**, 591–620.

Uccellini, L. W., 1976: Operational diagnostic applications of isentropic analysis. *Nat. Wea. Dig.*, **1**, 4–12.

——, D. R. Johnson and R. E. Schlesinger, 1979: An isentropic and sigma coordinate hybrid numerical model: Model development and some initial tests. *J. Atmos. Sci.*, **6**, 390–414.

University of Chicago, 1947: On the general circulation of the atmosphere in middle latitudes. *Bull. Amer. Meteor. Soc.*, **28**, 255–280.

Wexler, H., 1961: A boundary layer interpretation of the low-level jet. *Tellus*, **13**, 368–378.

Whittaker, T. M., and R. A. Petersen, 1977: Objective cross-sectional analysis incorporating thermal enhancement of the observed winds. *Mon. Wea. Rev.*, **105**, 147–153.

Young, J. A., 1973: A theory for isallobaric air flow in the planetary boundary layer. *J. Atmos. Sci.*, **30**, 1584–1592.

Paper 57

Reprinted from *Journal of the Atmospheric Sciences*, 36, pp. 390-414, 1979.

An Isentropic and Sigma Coordinate Hybrid Numerical Model: Model Development and Some Initial Tests

LOUIS W. UCCELLINI[1]

Space Science and Engineering Center[2]

DONALD R. JOHNSON

Space Science and Engineering Center and Department of Meteorology[2]

ROBERT E. SCHLESINGER

Department of Meteorology[2]

(Manuscript received 22 December 1977, in final form 15 November 1978)

ABSTRACT

Several recent efforts in the development of isentropic numerical models for atmospheric simulation have used sigma coordinates for prediction within the lower troposphere. Sigma coordinates are used in this region to provide a grid structure with uniform resolution for predicting planetary boundary layer processes and to avoid the problem of the intersection of information surfaces with the earth's surface. While successful simulations have been completed, none of the hybrid models provided for the conservation of mass and other properties with respect to transport through the common interface between the sigma and isentropic model domains. The primary problem is to match boundary conditions across the interface while providing for full interaction between the two model domains in the presence of diabatic heating and viscous forces without introducing spurious sources of mass, momentum or energy.

In this paper, a solution is presented which matches transport boundary conditions at the interface and conserves atmospheric properties without parameterization. The development of an isentropic and sigma coordinate hybrid model, based on the flux form of the primitive equations, is discussed. In this development phase, primary emphasis is on the unique modeling aspects related to the intersection of isentropic surfaces with the common boundary and the conservation of physical properties during exchange between the two model domains. Initial model simulations of a jet streak propagating in a zonal channel are presented that were designed to test the feasibility of the hybrid model approach. The results show that:

1) With respect to transport processes, mass, momentum and energy are conserved for the entire model domain. Vertical transport of properties through the interface between the two model domains is matched exactly.

2) The flux formulation of the hybrid model with a vertically staggered grid maintains a smooth transition at the interface level without need for artificial adjustments.

3) The time rates of change of total momentum and energy are associated primarily with physical forcing.

The effects of truncation errors resulting from initialization of emerging grid points in the truncated isentropic region, based on a redistribution of mass and momentum consistent with conservation of these properties, yields minor pressure and wind perturbations in the truncated and sigma grid volumes. Although the perturbations are stable and quickly damp, the need to initialize grid points in the sensitive model region above the boundary layer is a problem with hybrid models which merits further attention.

1. Introduction

In recent years, a renewed interest in utilizing isentropic coordinates for the study of midlatitude weather disturbances has been extended to modeling synoptic- and subsynoptic-scale phenomena within the isentropic framework (Eliassen and Raustein, 1968, 1970; Gall, 1972; Bleck, 1973, 1974; Shapiro, 1975, Deaven, 1976; Friend, 1976; Trevisan, 1976). An advantage of this approach is that for adiabatic flow isentropic surfaces become material surfaces. Thus, the depiction and prediction of three-dimensional transport processes are reduced to the two horizontal spatial dimensions (Starr, 1945; Charney and Phillips, 1953). This simplification suggests that truncation

[1] Present affiliation. NASA/Godcard Space Flight Center, Laboratory for Atmospheric Sciences, Code 914, Greenbelt, MD 20771.

[2] University of Wisconsin—Madison, Madison WI 53706.

0022-4928/79/030390-25$10.25

error in regions of adiabatic flow may be suppressed through the reduction in degrees of freedom of the advective terms in the basic equations. The need to estimate the vertical advection or transport in a hydrostatic atmosphere from a diagnostic evaluation of the divergent component of the motion field is eliminated. Another basic advantage is the variable spacing between isentropes which enhances the vertical resolution in baroclinic zones where strong wind shear exists. Therefore, upper tropospheric jet cores and attending frontal zones can be resolved even with the use of coarse-resolution horizontal grids (see Bleck, 1973; Shapiro, 1974).

The advantages of modeling in isentropic coordinates are offset by several deterrents. One of these is the behavior of isentropes near the ground. Although the intersection of isentropic surfaces with the earth's surface does not seem to be a serious obstacle, (Eliassen and Raustein, 1968, 1970; Bleck, 1974; Shapiro, 1975; Trevisan, 1976), the existence of superadiabatic and adiabatic layers, which occur most frequently in the boundary layer, cannot be tolerated. In addition, the tendency toward an adiabatic condition in the planetary boundary layer results in limited resolution in regions where vertical gradients and viscous stresses are large, and significant moisture, momentum and energy fluxes occur. The tendency for adiabatic conditions in the free atmosphere also results in limited vertical resolution in regions where vertical mixing through turbulent diffusion is likely enhanced by low static stability. Another degree of freedom introduced in any coordinate system with sloped upper and lower information surfaces (e.g., isentropic and sigma coordinates) is a mode of nonconvective flux of horizontal momentum and kinetic energy by pressure and viscous stresses on the upper and lower surfaces of a grid volume element (Johnson, 1977). These limitations of isentropic models may lead to deficiencies in weather forecasting, especially for predicting the boundary layer wind field and precipitation.

A proposed approach to resolve the boundary layer problem is a hydrid model (Gall, 1972; Deaven, 1976) which incorporates isentropic coordinates for the free atmosphere and sigma coordinates (Phillips, 1957) for the planetary boundary layer. One advantage of the hybrid approach is the optimum resolution that both coordinate systems offer for particular regions of the atmosphere. Second, superadiabatic and adiabatic conditions could exist in the boundary layer without need for adjustment due to the mathematical constraint that potential temperature must increase with height. The truncation of the isentropic domain above the planetary boundary layer also avoids the need to extrapolate values of the Montgomery streamfunction (ψ) to isentropic surfaces "beneath the ground" in order to calculate the pressure gradient force in the equations of motion. Values of ψ on isentropic surfaces beneath the interface separating the two model domains can be determined by interpolation from the atmospheric structure within the sigma-coordinate boundary layer. This is an important feature in estimating the horizontal pressure-gradient force near the lower boundary of an isentropic model domain.

Among the efforts to develop a hybrid model, Gall (1972) combined an "inactive advective" planetary boundary layer in sigma coordinates with isentropic coordinates for the free atmosphere to model a propagating jet streak. Deaven (1976) joined the isentropic representation of the free atmosphere with an active planetary boundary layer in sigma coordinates in a two-dimensional model to study dynamically and thermodynamically forced circulations. His results were realistic for a variety of two-dimensional circulations. However, neither Gall's nor Deaven's models possessed degrees of freedom for transport of mass or other properties through the common interface separating the sigma and isentropic model domains. A resulting limitation would occur for the prediction of deep tropospheric mass circulations characterized by inflow in the boundary layer, upward transport throughout the entire vertical extent of the troposphere and outflow in the high troposphere. In this situation the material surface at the interface would eventually span the entire troposphere over the region of ascent and be severely deformed. The determination of pressure and viscous forces near or on the interface which satisfy conservation principles would be compromised by the increased slope of the sigma surfaces. Recently, Friend *et al.* (1977) completed a three-dimensional hybrid model which included an active boundary layer and vertical advection through the interface. However, some uncertainty exists concerning the accuracy and success of their methods. The model does not insure that the solutions for the predictive equations at the interface and on isentropic surfaces immediately above the interface are consistent and satisfy transport constraints. The separation of the solutions near the interface thus requires an artificial adjustment to smooth the mass and wind fields across the interface. These adjustments, however, violate conservation principles and appear to generate spurious gravity waves (Friend, 1976).

The primary problem with the hydrid model approach has been to match boundary conditions across the interface which provide for full interaction of the two model domains in the presence of diabatic heating and viscous forces, without introducing spurious sources of mass, momentum or energy. The purposes of this paper are as follows:

1) to pose a solution for matching boundary conditions across the interface of an isentropic and sigma coordinate hydrid model;

2) to design a hydrid model based on the flux form of the primitive equations that allows direct vertical exchange between the model domains, satisfies con-

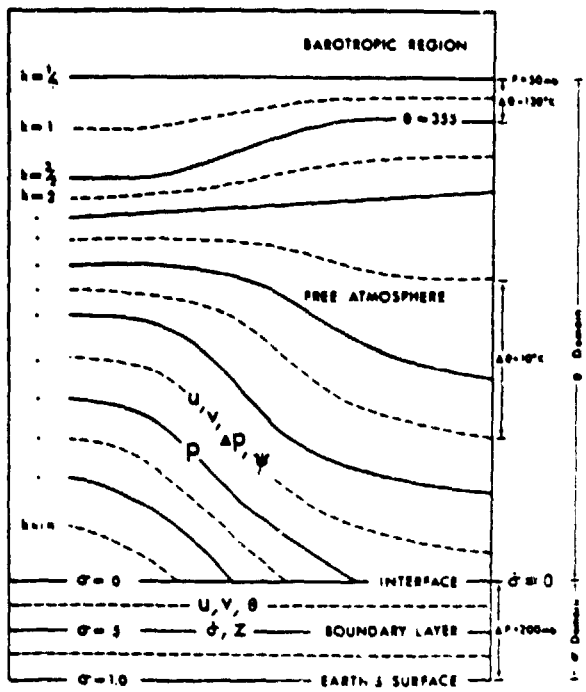


FIG. 1. Cross section of hybrid model showing barotropic, isentropic free atmosphere and sigma boundary-layer domains. Solid and dashed surfaces indicate vertical staggering of variables. n is the number of isentropic layers.

servation principles with respect to transport processes, and maintains a smooth transition across the interface without need for artificial adjustment or parameterization schemes;

3) to present initial model results of experiments designed to test the feasibility of the hybrid model approach.

A solution for the problem of matching transport across the interface while conserving atmospheric properties is presented in Section 2. A description of a three-dimensional hydrid model is discussed in Section 3. In this development phase, primary emphasis is on the unique modeling aspects related to the intersection of the isentropic surfaces with the common boundary and the conservation of physical properties during exchange between the two model domains. A detailed description of the modeling techniques can be found in the Scientific Report by Uccellini *et al.* (1977). The results from the initial model experiments are presented in Section 4 with a summary and future research plans presented in Section 5.

2. Budget integrals for the hybrid model and conservation of properties during exchange across the common interface

The isentropic and sigma model domains are illustrated in Fig. 1. Their interface is the common boundary through which transport from one domain must equal the transport into the other domain. The following development focuses on matching the isentropic and sigma coordinate boundary conditions at the interface.

The budget integrals of an arbitrary physical property (f) for the isentropic (F_2) and sigma (F_1) portions of the hybrid model, with the common boundary σ_B are

$$F_2 = \int_A \int_{\theta_B}^{\theta_T} \rho J_\theta f \, d\theta \, dA, \tag{1}$$

$$F_2 = -\int_A \int_1^{\sigma_B} \rho J_\sigma f \, d\sigma \, dA, \tag{2}$$

where all limits of integration are invariant except the isentropic budget's lower limit θ_B. With the Jacobian defined by absolute value, the negative sign in (2) is required for F_1 to be positive since σ is a monotonic decreasing function with height. A complete list of symbols is presented in Appendix A.

By Leibniz's rule for the differentiation of integrals, the time rates of change of (1) and (2) are

$$\frac{dF_2}{dt} = \int_A \int_{\theta_B}^{\theta_T} \frac{\partial}{\partial t}(\rho J_\theta f) \, d\theta \, dA - \int_A \rho J_\theta f \frac{d\theta_B}{dt} dA \Big|_{\theta_B(x,y,\sigma_B,t)}, \tag{3}$$

$$\frac{dF_1}{dt} = -\int_A \int_1^{\sigma_B} \frac{\partial}{\partial t_\sigma}(\rho J_\sigma f) \, d\sigma \, dA. \tag{4}$$

The last term in (3) represents the changes associated with the expansion or contraction of the isentropic domain due to time variations of θ_B over the surface of integration. This surface of integration is the common boundary on which distributions of variables are specified by the intersection of σ_B with the isentropic coordinate structure.

The use of a generalized transport equation (Johnson, 1977)

$$\frac{\partial}{\partial t_\eta}(\rho J_\eta f) + \nabla_\eta \cdot (\rho J_\eta f \mathbf{U}) + \frac{\partial}{\partial \eta}\left(\rho J_\eta f \frac{d\eta}{dt}\right) = \rho J_\eta \frac{df}{dt}, \tag{5}$$

in (3) and (4) and integration yields

$$\frac{dF_2}{dt} = \int_A \int_{\theta_B}^{\theta_T} \rho J_\theta \frac{df}{dt} d\theta \, dA - \oint \int_{\theta_B}^{\theta_T} \rho J_\theta f \mathbf{U} \cdot \mathbf{n} \, d\theta \, dL + \int_A \rho J_\theta f \left(\frac{d\theta}{dt} - \frac{d\theta_B}{dt}\right)_{\theta_B(x,y,\sigma_B,t)} dA, \tag{6}$$

$$\frac{dF_1}{dt} = -\int_A \int_1^{\sigma_B} \rho J_\sigma \frac{df}{dt} d\sigma \, dA + \oint \int_1^{\sigma_B} \rho J_\sigma f \mathbf{U} \cdot \mathbf{n} \, d\sigma \, dL - \int_A \rho J_\sigma f \frac{d\sigma}{dt}_{\sigma_B} dA, \tag{7}$$

where the vertical mass transports $\rho J_\theta d\theta/dt$ and $\rho J_\sigma d\sigma/dt$ vanish at the top and bottom of the atmosphere, respectively. The last term in (6) is the transport relative to the coordinate θ_B, while the last term in (7) is the transport of the property f through the boundary σ_B. By the chain rule, the time variation of $\theta_B(x,y,\sigma_B,t)$ on the boundary σ_B is

$$\frac{d\theta_B}{dt}=\frac{\partial\theta_B}{\partial t_\sigma}+\mathbf{U}\cdot\nabla_\sigma\theta_B. \tag{8}$$

The substitution of (8) into (6) yields

$$\frac{dF_2}{dt}=\int_A\int_{\theta_B}^{\theta_T}\rho J_\sigma\frac{df}{dt}d\theta dA-\int_{\theta_B}^{\theta_T}\oint\rho J_\theta f\mathbf{U}\cdot\mathbf{n}dLd\theta$$
$$+\int_A\rho J_\theta f\left[\frac{d\theta}{dt}-\left(\frac{\partial\theta}{\partial t_\sigma}+\mathbf{U}\cdot\nabla_\sigma\theta\right)\right]\Bigg|^{\sigma_B}dA, \tag{9}$$

where the subscripts for θ_B in (9) have been deleted since its distribution is specified by the intersection of the surface of integration σ_B with the isentropic domain. The final result for (9), accomplished with the expansion of $d\theta/dt$ and the Jacobian relation

$$J_\theta=-J_\sigma\frac{\partial\sigma}{\partial\theta}, \tag{10}$$

is

$$\frac{dF_2}{dt}=\int_A\int_{\theta_B}^{\theta_T}\rho J_\sigma\frac{df}{dt}d\theta dA-\int_{\theta_B}^{\theta_T}\oint\rho J_\theta f\mathbf{U}\cdot\mathbf{n}dLd\theta$$
$$-\int_A\rho J_\sigma f\frac{d\sigma}{dt}\Bigg|^{\sigma_B}dA. \tag{11}$$

This result verifies that in the absence of internal sources or sinks of properties, the time rates of change of the budget integrals for the isentropic and sigma domains are determined by transport across boundaries and that the transport relative to the coordinate θ_B in (6) is identical to the transport of the property f through the boundary σ_B. With identical expressions of opposite signs in (7) and (11) for the transport of any property across the common boundary of the hybrid isentropic-sigma model, integrals of properties determined from a finite-difference flux formulation will be conservative.

3. The numerical model

The description of the numerical mode[illegible] to test the feasibility of the hydrid approach is [illegible]ted in this section. The discussion focuses on t[illegible]tions for both model domains, the finite-difference methods for computing transport, the calculation of the Montgomery streamfunction and its gradient near the interface for truncated isentropic grid volumes, boundary conditions and initialization procedures.

a. *Model equations*

The finite difference flux forms of the primitive equations for isentropic and sigma coordinates in Tables I and II, respectively, stem from a generalized transport equation for basic physical properties. The total property within each grid volume ρJf is equal to $\rho^I J_\theta f$ for isentropic coordinates and $\rho J_\sigma J_\sigma f$ for sigma coordinates, where $J_\theta=\partial z/\partial\theta$ and $J_\sigma=\partial z/\partial\sigma$ transform the vertical coordinate and $J_\sigma(x,y,\Lambda,\Phi)$ transforms

TABLE I. Flux form of hydrostatic primitive equations for modified spherical grid: Isentropic coordinates.

Mass continuity equation

$$\frac{\partial}{\partial t}(\rho J_\theta J_\sigma)=-\frac{1}{a}\left[\frac{\partial}{\partial\Lambda}(\sigma^2\rho J_\theta u)+\frac{\partial}{\partial\Phi}(\rho J_\theta J_\sigma v)\right]-\mathrm{VF} \tag{I1}$$

normal: $\mathrm{VF}=J_\sigma(\dot\theta\rho J_\theta|^{\theta_T}-\dot\theta\rho J_\theta|^{\theta_L})$

truncated: $\mathrm{VF}=J_\sigma(\dot\theta\rho J_\theta|^{\theta_T}+\dot\sigma\rho J_\sigma|^{\sigma_B})$

u momentum equation

$$\frac{\partial}{\partial t}(\rho J_\theta J_\sigma u)=-\frac{1}{a}\left[\frac{\partial}{\partial\Lambda}(\sigma^2\rho J_\theta uu)+\frac{\partial}{\partial\Phi}(\rho J_\theta J_\sigma uv)\right]-\mathrm{VF}+\rho J_\theta J_\sigma\left(fv-\frac{\cos\phi_0}{a\cos\phi}\frac{\partial\psi}{\partial\Lambda}+F_u\right) \tag{I2}$$

normal: $\mathrm{VF}=J_\sigma(\dot\theta\rho J_\theta u|^{\theta_T}-\dot\theta\rho J_\theta u|^{\theta_L})$

truncated: $\mathrm{VF}=J_\sigma(\dot\theta\rho J_\theta u|^{\theta_T}+\dot\sigma\rho J_\sigma u|^{\sigma_B})$

v momentum equation

$$\frac{\partial}{\partial t}(\rho J_\theta J_\sigma v)=-\frac{1}{a}\left[\frac{\partial}{\partial\Lambda}(\sigma^2\rho J_\theta uv)+\frac{\partial}{\partial\Phi}(\rho J_\theta J_\sigma vv)\right]-\mathrm{VF}+\rho J_\theta J_\sigma\left(-fu-\frac{1}{a}\frac{\partial\psi}{\partial\Phi}+F_v\right) \tag{I3}$$

normal: $\mathrm{VF}=J_\sigma(\dot\theta\rho J_\theta v|^{\theta_T}-\dot\theta\rho J_\theta v|^{\theta_L})$

truncated: $\mathrm{VF}=J_\sigma(\dot\theta\rho J_\theta v|^{\theta_T}+\dot\sigma\rho J_\sigma v|^{\sigma_B})$

Exner function $\pi=c_p(p/p_{00})^\kappa$ (I4)

Montgomery stream function $\psi=\phi_T+gz$

Hydrostatic relation $\psi_\theta=\psi_B+\int_{\theta_B}^{\theta}\pi d\theta$ (I5)

First law of thermodynamics $\dot\theta=\pi^{-1}Q_m$ (I6)

TABLE II. As in Table I except for sigma coordinates.

Mass continuity equation

$$\frac{\partial}{\partial t}(\rho J_\sigma J_\sigma) = -\frac{1}{a}\left[\frac{\partial}{\partial \Lambda}(a^2 \rho J_\sigma u) + \frac{\partial}{\partial \Phi}(\rho J_\sigma J_\sigma v)\right] - J_\sigma \frac{\partial}{\partial \sigma}(\rho J_\sigma \dot{\sigma}) \tag{II1}$$

u momentum equation

$$\frac{\partial}{\partial t}(\rho J_\sigma J_\sigma u) = -\frac{1}{a}\left[\frac{\partial}{\partial \Lambda}(a^2 \rho J_\sigma uu) + \frac{\partial}{\partial \Phi}(\rho J_\sigma J_\sigma uv)\right] - J_\sigma \frac{\partial}{\partial \sigma}(\rho J_\sigma u\dot{\sigma}) \quad J_\sigma J_\sigma\left[fv - \frac{\cos\phi_0}{a\cos\phi}\left(\theta\frac{\partial \pi}{\partial \Lambda} + g\frac{\partial z}{\partial \Lambda}\right) + F_u\right] \tag{II2}$$

v momentum equation

$$\frac{\partial}{\partial t}(\rho J_\sigma J_\sigma v) = -\frac{1}{a}\left[\frac{\partial}{\partial \Lambda}(a^2 \rho J_\sigma vu) + \frac{\partial}{\partial \Phi}(\rho J_\sigma J_\sigma vv)\right] - J_\sigma \frac{\partial}{\partial \sigma}(\rho J_\sigma v\dot{\sigma}) + J_\sigma J_\sigma\left[-fu - \frac{1}{a}\left(\theta\frac{\partial \pi}{\partial \Phi} + g\frac{\partial z}{\partial \Phi}\right) + F_v\right] \tag{II3}$$

First law of thermodynamics

$$\frac{\partial}{\partial t}(\rho J_\sigma J_\sigma \theta) = \rho J_\sigma J_\sigma \dot{\theta} - \frac{1}{a}\left[\frac{\partial}{\partial \Lambda}(a^2 \rho J_\sigma \theta u) + \frac{\partial}{\partial \Phi}(\rho J_\sigma J_\sigma \theta v)\right] - J_\sigma \frac{\partial}{\partial \sigma}(\rho J_\sigma \theta\dot{\sigma}) \tag{II4}$$

Exner function

$$\pi = c_p(p/p_{00})^\kappa \tag{II5}$$

Hydrostatic relation

$$z)_\sigma = z)_{\sigma=1} + \frac{R p^*}{g c_p}\int_\sigma^1 \frac{\theta\pi}{p}\,d\sigma \tag{II6}$$

Definition of sigma

$$\sigma = \frac{p - p_B}{p_s - p_B} = \frac{p - p_B}{p^*}, \quad 0 \leq \sigma \leq 1, \quad p_B \leq p \leq p_s \tag{II7}$$

horizontal area in Cartesian coordinates to modified spherical coordinates. The modified spherical coordinate, with J_σ equal to $a^2\cos\phi/\cos 45°$, provides for a square grid area at 45° rather than at the equator while retaining the poleward convergence of the meridians and therefore reduces the grid distortion within the model domain (Uccellini *et al.*, 1977). Since no information exists for vertical variation of the mass (ρJ) or specific properties (f) within grid volumes, each quantity is represented by its vertical average. With the hydrostatic assumption, the total property within the vertically integrated grid volume becomes $(\Delta p\ g)J_\sigma f$ for both domains. Here Δp is the absolute value of the difference between the pressures on the lower and upper surfaces of the grid volume, regardless of whether the grid volume is complete with top and bottom isentropic surfaces or is truncated by the interface separating the model domain. In the sigma domain Δp is equal to $p^*/2$. In Eqs. (I1)–(I3) the vertical fluxes (VF) are integrated over the vertical extent of normal and truncated volumes.

Gall (1972) conducted experiments with both adiabatic and diabatic conditions. In his diabatic experiments, he predicted the vertical transport $\rho J_\sigma d\theta/dt$ through the use of a water vapor transport equation. In our initial attempts with the hybrid model, adiabatic conditions are imposed to study mass conservation. With this constraint, the layer-average hydrostatic mass distribution and the average pressure $\bar{p}(\theta)$ must be conserved in the isentropic domain since the isentropic vertical transport vanishes. Since the tendency of the area-average pressure distribution over isentropic surfaces that extend into the sigma domain must also vanish, the time variation of $\bar{p}(\theta)$ is a diagnostic measure of conservation of the mass distribution above any isentropic surface within the entire model domain.

Another simplification is introduced by defining the vertical extent of the sigma-coordinate boundary layer to be an increment of mass corresponding to 200 mb. With this simplification p^* reduces to a constant and the vertical motion σ, determined by the integration of the equation of continuity (II1), is

$$\sigma = J_\sigma^{-1}\int_\sigma^{\sigma=1}\frac{1}{a}\left[\frac{\partial}{\partial \Lambda}(a^2 u) + \frac{\partial}{\partial \Phi}(J_\sigma v)\right]d\sigma, \tag{12}$$

where the boundary condition that σ vanishes at the earth's surface is applied. With the diagnostic evaluation of σ, the vertical transport of any physical property (e.g., mass, momentum and energy) through the common boundary is determined without the need for parameterization. The vertical transport at the interface, which is computed from information in the sigma domain and on the interface, constitutes the lower boundary information needed for the truncated grid volumes in the isentropic domain. With these simplifications the system of equations is complete for numerical integration.

b. Grid structure

Since the experiment is designed to simulate a propagating jet streak, the model is applied to a zonal channel ~30° wide centered at 45°N. This channel is bounded to the north and south by impervious free-slip walls, while cyclic continuity conditions are assumed on the longitudinal boundaries. Calculations are made for only one wavelength.

The vertical structure of the model (illustrated in Fig. 1) is divided into three regions. The sigma coordinate boundary layer, with a depth of 200 mb, extends from the earth's surface ($\sigma = 1$) to the interface ($\sigma = 0$). The isentropic domain for the free atmosphere

extends from the interface to the 475 K isentropic surface which is assumed to have a uniform pressure of 50 mb. With the pressure assumed to be uniform on the 475 K surface, the static stability measure $\partial p/\partial\theta$ is horizontally invariant for the layer from 475 K to the top of the atmosphere.

The grids of the isentropic and sigma model domains are vertically staggered to simplify the finite differencing of the flux form of the equations. Solid lines in the isentropic domain are information surfaces for pressure, while the dashed lines are surfaces for u, v, Δp, ψ and θ. In the sigma domain solid lines below the interface are information surfaces for σ and z. The dashed lines are surfaces for u, v, θ and π. On the interface $\dot{\sigma}$, z, u, v, θ, ψ and p must be specified. For these experiments, the vertical spacing in the isentropic domain below 355 K is 10 K, while above this level it is 120 K. The vertical increment of sigma is 0.5. The horizontal grid resolution is 275 km.

c. Determining interface parameters and surface pressure

The grid-point values of θ, u and v on the interface are not predicted directly but are interpolated to that level after the predictions of Δp, u, v and θ, u, v are completed in the isentropic and sigma domains, respectively. The pressure on all "solid" isentropic surfaces below the 475 K surface *and on the interface* is obtained directly by the downward vertical integration of $\partial p/\partial\theta$ which is predicted from (11). Since the prediction of $\partial p/\partial\theta$ for the grid volume immediately above the interface includes the mass flux through the interface ($\rho J_\theta \sigma_B$), the pressure tendency on the interface is a function of the net horizontal mass divergence within the isentropic domain and the net vertical mass flux from the sigma model domain. Since σ is determined as a diagnostic variable with the constraint of constant p^*, the vertical mass flux through the interface is determined from the net horizontal mass divergence between the earth's surface and the interface and thus acts to maintain constant mass between sigma surfaces. Constant p^* also implies that the pressure tendency at any sigma level including the earth's surface is equal to the interface pressure tendency. Note that by partial differentiation of (II7), $\partial p_s/\partial t_\sigma = \partial p_B/\partial t_\sigma$ with a constant p^* and the pressure tendency within the sigma domain is independent of the vertical coordinate. Therefore, the surface pressure always remains 200 mb greater than p_B.

Once p_B is determined, a third-order interpolation of θ with respect to p from the 0.25 and 0.75 sigma surfaces and the first two solid isentropic surfaces above the interface (Fig. 1) is utilized to specify θ_B. The wind components are then linearly interpolated to the interface from the 0.25 sigma surface and the first dashed isentropic surface above the interface with respect to θ. Since the vertical motion $\dot{\sigma}_B$ through the interface is determined through a direct evaluation of (12), these methods provide for a smooth transition and unique determination of mass and momentum between the model domains, preclude spurious sources of the properties and processes on and near the interface, and allow for a direct calculation of the vertical flux without need for parameterization or special adjustment procedures.

d. Finite differencing: Spatial

The box method of spatial differencing (Kurihara and Holloway, 1967) was chosen for this model since the method conserves integral quantities of mass, momentum and energy with respect to transport processes. The application of this method is relatively simple for the free-isentropic and sigma domains where the lateral walls of the box are located halfway between the center and the four surrounding grid points, and the top and bottom of the box are "solid" coordinate surfaces (Fig. 2). The flux of any physical quantity through the lateral walls is determined by calculating an average mass transport vector and an average value for a property at the wall using grid-point data on either side of the wall. For example, the flux divergence of the physical property f in the longitudinal direction is

$$\frac{\frac{1}{2}(\Delta p_1 u_1 + \Delta p_0 u_0)\frac{1}{2}(f_1+f_0) - \frac{1}{2}(\Delta p_2 u_2 + \Delta p_0 u_0)\frac{1}{2}(f_2+f_0)}{\Delta x} \tag{13}$$

The flux through the top and bottom of each box in the sigma domain is determined by using the product of $\dot{\sigma}$, determined at the solid surface, and the average

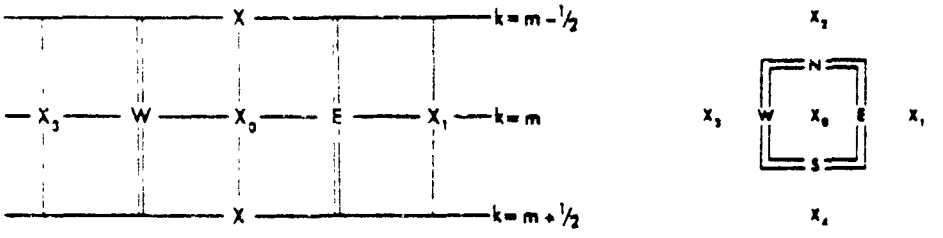


FIG. 2. Cross section of grid box (left) used in finite-difference scheme and plan view (right) of box on dashed coordinate surface. X is position of grid points and doubled lines are lateral walls of grid box with directions indicated.

values of the property f, calculated from values at grid points immediately above and below the solid surfaces. Since the flux from one box is exactly equal to the flux into the adjacent box, interior fluxes cancel and net changes by transport are due solely to fluxes through the boundaries of the model domain. With impervious lateral boundaries and cyclic conditions in the longitudinal direction, conservation of integral properties with respect to transport is assured.

In the lower portion of the free atmosphere the box method is complicated by truncated isentropic grid volumes. A "truncated" volume exists for the lowest dashed surface grid point in each grid column in the isentropic domain, and for any volume where the lower solid isentropic surface intersects the interface within the $2\Delta x$ or $2\Delta y$ interval. (See Fig. 15 in Appendix B where the application of the box method for truncated grid volumes is discussed.)

e. Computation of horizontal gradients of the Montgomery streamfunction

The Montgomery streamfunction is determined from a vertical integration of the hydrostatic equation (I5) using mean pressure for the integration interval. In truncated grid volumes, pressure is interpolated to the midpoint of the integration interval. The net internal source of momentum related to the Coriolis and pressure gradient forces is evaluated at the central grid point. The gradient of ψ is approximated by a standard centered-space difference in cases where the dashed isentropic surfaces do not intersect the interface within a $2\Delta x$ or $2\Delta y$ grid interval. For surfaces which intersect the interface, the calculation of $\nabla_\theta\psi$ is complicated. With the hybrid model approach, values of ψ on isentropic surfaces beneath the interface may be determined by interpolation from the atmospheric structure within the sigma coordinate boundary layer, rather than by extrapolations of ψ onto "underground" isentropic surfaces. Further details of the interpolation procedure are discussed by Uccellini *et al.* (1977).

f. Finite difference: Time

The leapfrog scheme combined with a time filter developed by Asselin (1972) was used in this model. Bleck (1974) and Haltiner and McCollough (1975) demonstrated that this scheme minimizes time splitting and produces results similar to those obtained by the more stable Euler backward method, but at nearly half the cost. To further reduce the spurious perturbations associated with the leapfrog scheme, the standard time increment ($\Delta t = 300$ s) is initially divided by 2^7. This modified time increment is then doubled the first seven time steps, keeping the initial field as the first time level. Once the standard Δt is reached and the predictions for $t = 600$ s completed, the doubling procedure ceases, and subsequent predictions are carried out in the normal fashion. The properties of the filtered leapfrog scheme are examined by a one-dimensional linear analysis of the advection equation by Uccellini, *et al.* (1977).

g. Diffusion

The large-scale horizontal diffusion is estimated from the Laplacian of the generalized property (f) by the Dufort-Frankel scheme (Richtmyer and Morton, 1967)

$$\nu\nabla^2 f = \frac{\nu[f^{\tau}_{i,j+1} + f^{\tau}_{i,j-1} + f^{\tau}_{i+1,j} + f^{\tau}_{i-1,j} - 2(f^{\tau+1}_{i,j} + \bar{f}^{\tau-1}_{i,j})]}{h^2}, \tag{14}$$

which has been modified to include the filtered variable $\bar{f}^{\tau-1}_{i,j}$ (Bleck, 1974), and where

$$\nu = Kh^2|D|. \tag{15}$$

The values of K are specified for the model experiments in Table III (see Appendix C). When the isentropic surface intersects the interface within the $2\Delta x$ or $2\Delta y$ interval, the value of f at the submerged grid point is obtained by linear interpolation within the boundary layer. If the submerged isentropic surface also intersects the lowest sigma surface within the $2\Delta x$ or $2\Delta y$ interval, the value of f on the lowest sigma surface is assigned as the grid-point value. The characteristics of the Dufort-Frankel scheme, including the effect of the time filter and space truncation errors, are discussed by Uccellini *et al.* (1977).

h. Boundary conditions

Lateral boundary conditions for this model are identical to the physical conditions used in Gall's model (1972). Cyclic continuity is assumed for the longitudinal boundaries. The latitudinal boundaries are assumed to be midway between an exterior and first interior row of grid points. The free-slip condition is satisfied by requiring the u component at the exterior and first interior grid point to be equal. The condition of impervious walls requires that v, $\partial v/\partial t$, $\partial v/\partial x$ and $\partial v/\partial\theta$ must be zero. For v to be zero on the boundary, v on the exterior row must be equal to $-v$ at each grid point on the first interior row. As a result, the v equation of motion on the boundary reduces to the geostrophic relation

$$fu_g = \frac{\partial\psi}{\partial y}. \tag{16}$$

Since flow along the wall remains geostrophic, ψ on the exterior rows is determined from the value of u_g on the wall and ψ on the first interior grid point.

Including diffusion terms requires additional conditions to be imposed on the walls to prevent diffusion across boundaries from changing the integral quantities of momentum and energy (Smagorinsky, 1958). This

is accomplished by setting all other variables on the exterior row equal to their values on the first interior row.

The same conditions are used in the sigma domain. The initial value of $\partial \pi_B / \partial y$, set equal to zero on the lateral boundaries of the sigma domain, remains time independent from the antisymmetric conditions assumed for the meridional mass transport $v\ \partial p / \partial \theta$ on the exterior and first interior grid points. With this condition, the v equation of motion on the boundary reduces to

$$f u_g = -g \frac{\partial z}{\partial y}. \tag{17}$$

The height z on the exterior rows is determined in the same manner as ψ is in the isentropic domain.

The upper boundary conditions for the model are that the pressure on the 475 K surface is time independent and the horizontal divergence of the adiabatic mass transport is assumed to be zero. With these conditions, the tendency of $\partial p / \partial \theta$ vanishes and the upper region remains barotropic. As a consequence of barotropy and nondivergence of mass, the vorticity within the upper region is conserved and no explicit calculations are needed (Gall, 1972).

These boundary conditions insure that no net source of mass, momentum or energy occurs from transport through the boundaries. Any net source detected in the model experiments must be related to physical processes or numerical deficiencies within the model domain.

4. Initial model results

The initial hybrid model simulations of a jet streak propagating in a zonal channel are utilized to test the feasibility of the hybrid model approach. The specific

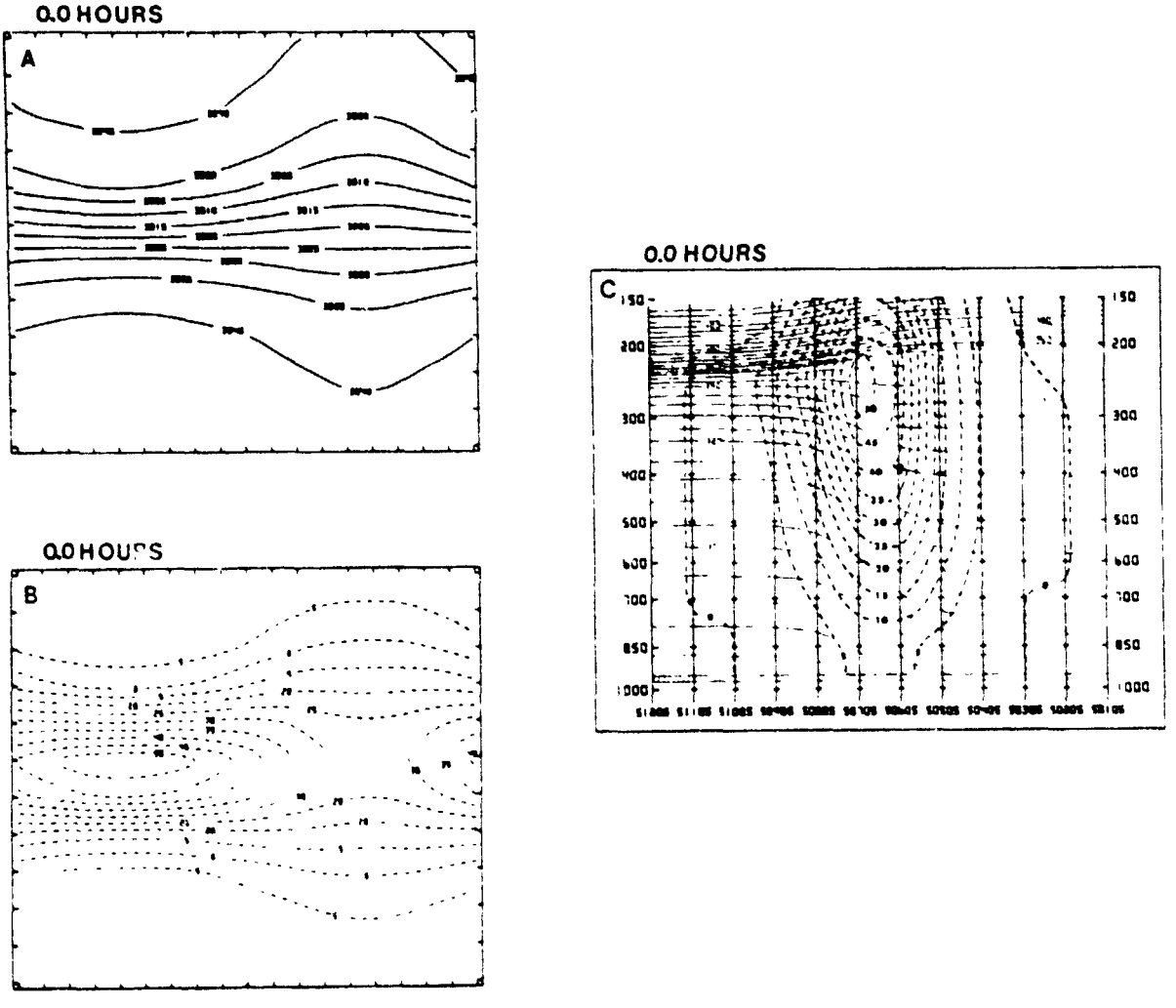


FIG. 3. (A) Initialized ψ ($\times 10^6$ m^2 s^{-1}) and (B) geostrophic isotachs (m s^{-1}) for 340 K isentropic surface. (C) Vertical cross section through jet maximum along column 5 in B), solid lines are isentropes (K) and dashed lines are isotachs (m s^{-1}).

purpose of the model experiments is to demonstrate that:

- The physical properties are conserved for the entire model domain.
- The vertical exchange of physical properties through the interface does not result in spurious sources of mass, momentum or energy.
- A smooth transition of the mass and wind fields across the interface is maintained.
- The predictive equations remain stable in the truncated isentropic domain even as isentropic surfaces rise and sink through the interface.

The assurance that numerical solutions exist for these problems will be established before the hybrid approach is applied to studies of adjustment, frontogenesis and cyclone development.

The initialization of the model is described in Appendix C. The initial conditions were chosen to be identical with Gall's (1972) model experiments for comparison purposes and to assess the impact of adding an active sigma-coordinate boundary layer to the isentropic model. After a brief description of model simulations of a propagating jet streak, the model's ability to conserve mass and to maintain a smooth transition between the two model domains is discussed. The model's response to the initialization procedure for grid points which emerge into the truncated isentropic domain is also discussed.

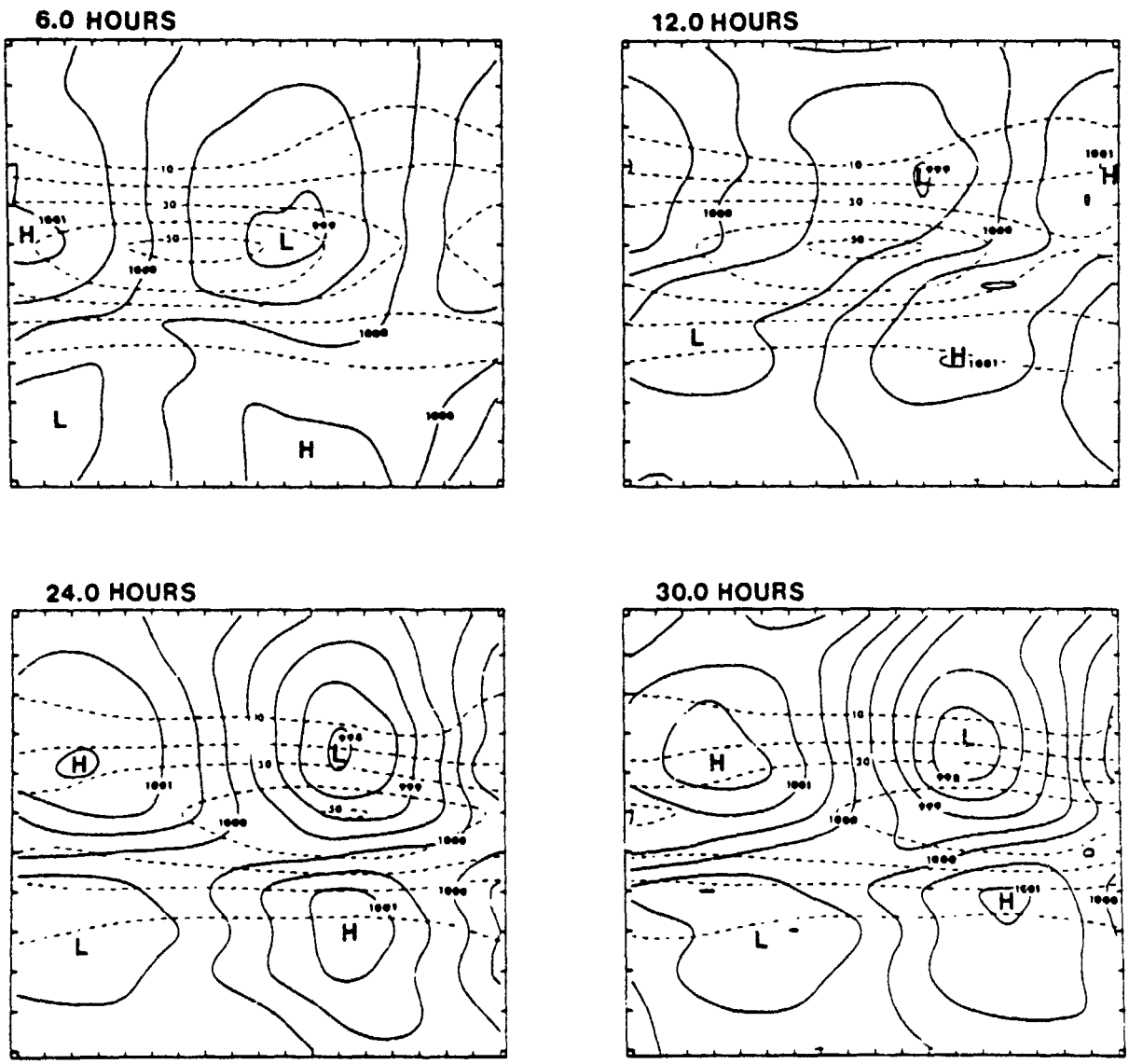


FIG. 4. Isotachs (dashed lines, m s^{-1}) for 340 K isentropic surface, and surface pressure (solid lines, mb) for model Experiment E1.

a. Description of the model simulation of a propagating jet streak

The results from the two 36 h adiabatic simulations, one without diffusion (E1) and the other with diffusion (E2), are presented in this section. The initial conditions, described in Appendix C, are exactly the same for both experiments (see Table III in Appendix C for the values assigned to the initialization parameters). The initial wave and jet streak are illustrated in Fig. 3 by the ψ and isotach analyses on the 340 K surface and by the vertical cross section through the region of maximum wind. The vertical cross section in Fig. 3 is completed by constructing soundings at each grid point and by using an objective cross section program (Whittaker and Petersen, 1977) with the model u wind component utilized for the isotach analysis.

During Experiment E1, the jet streak propagated eastward at rates which diminished from an average of 30 m s^{-1} for the first 12 h period, to 24 m s^{-1} for the second 12 h period and to 18 m s^{-1} for the third 12 h period, as the maximum wind in the core decreased from an initial speed of 54.8 to 48.0 m s^{-1} during the 36 h. For jet streaks with maximum winds exceeding the propagation rate of the streak, direct and indirect transverse mass circulations develop in the entrance (confluent) and exit (diffluent) regions, respectively (University of Chicago, 1947; Namias and Clapp, 1949; Bjerknes, 1951; Riehl *et al.*, 1952; Murray and Daniels, 1953; Sawyer, 1956; Newton, 1959; Eliassen,

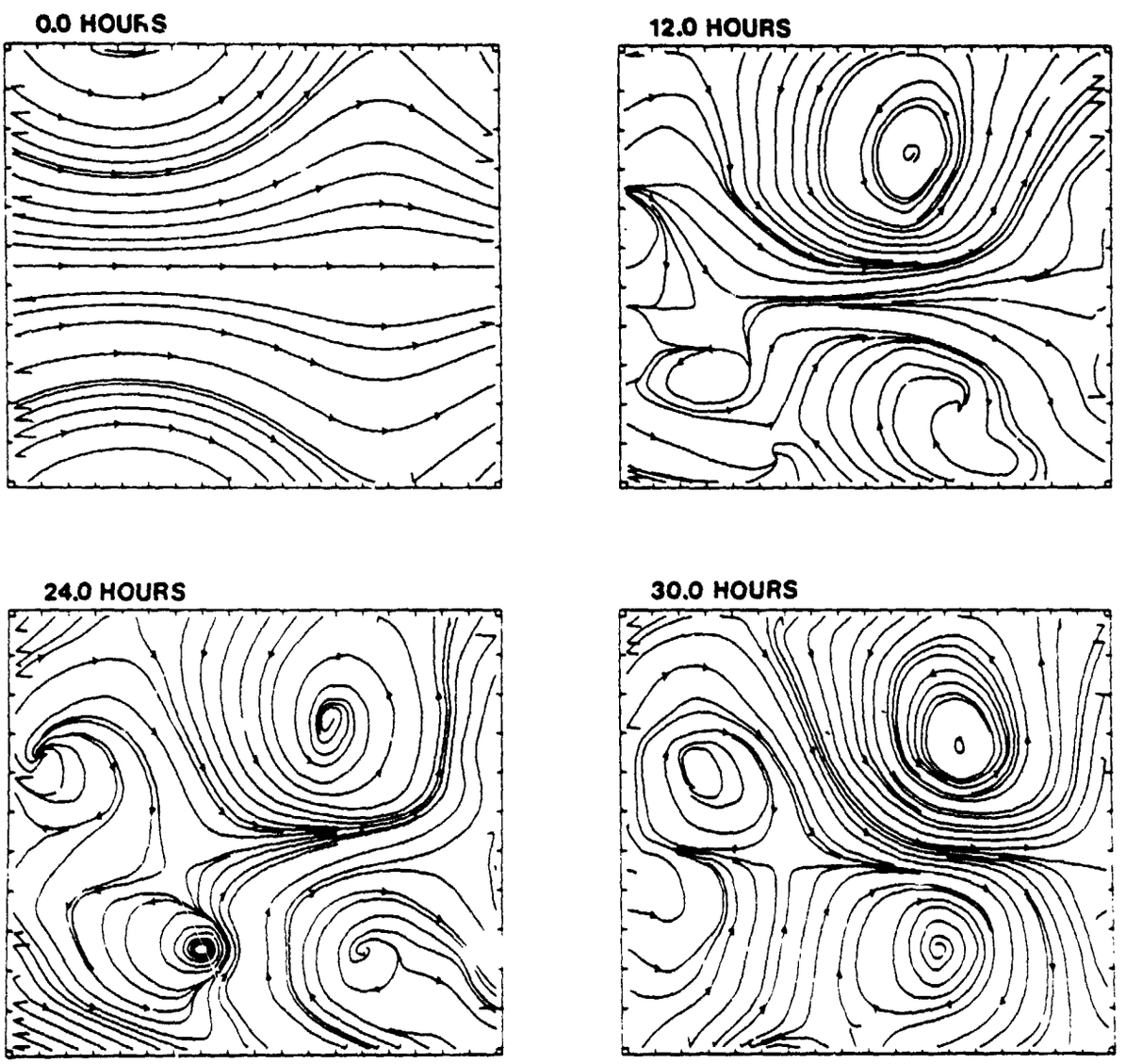


FIG. 5. Surface streamlines for model Experiment E1, analyzed for the model domain up to the first interior row of grid points along the north and south boundaries.

1962; Reiter, 1963, 1967, 1969; Cahir, 1971; Gall, 1972; and Hoskins and Bretherton, 1972 among many others). A manifestation of the transverse circulations is the four-cell pressure pattern which developed as the jet propagated eastward (Fig. 4), a pattern which should be expected from the quasi-geostrophic adjustments in the entrance and exit regions of the jet streak and which favorably compares with Gall's model results. As described by Reiter (1967) and by Gall, the net mass transport from the cyclonic to anticyclonic side of the streak by the indirect circulation decreases the surface pressure on the cyclonic side and increases the pressure on the anticyclonic side in the exit region. In the entrance region the reverse is true. The net mass transport by the direct circulation increases the surface pressure on the cyclonic side and decreases the pressure on the anticyclonic side of the jet. The mutual adjustment of the low-level wind field to the surface pressure pattern is illustrated by streamline analyses for the 0.75 sigma surface (Fig. 5). By 12 h, the initial wave has transformed into the basic four-cell pattern with cyclonic flow associated with low pressure and anticyclonic flow with high pressure that persisted throughout the remaining 24 h of integration.

Figs. 4 and 5 illustrate an apparent decoupling between the upper tropospheric jet streak and the surface pressure field by 30 h as the streak continued propagating eastward at a faster rate than the surface low in the exit region. During the last 18 h of the model simulation, the along-stream variation of the u wind component at the jet streak level decreased from near 20 to 6 m s^{-1} over the position of the surface low located beneath the exit region of the streak. Since this along-stream variation of the u component is a primary forcing function for the cross-stream mass transports (Bjerknes, 1951), the overall development of the surface pressure field should gradually diminish as the jet streak weakens. The coupling of the upper and lower tropospheric circulations also weakens through the reduction of the cross-stream mass transport.

The vertical velocity field ($\dot{\sigma}$) which developed at the interface during the first 15 h of the model simulation

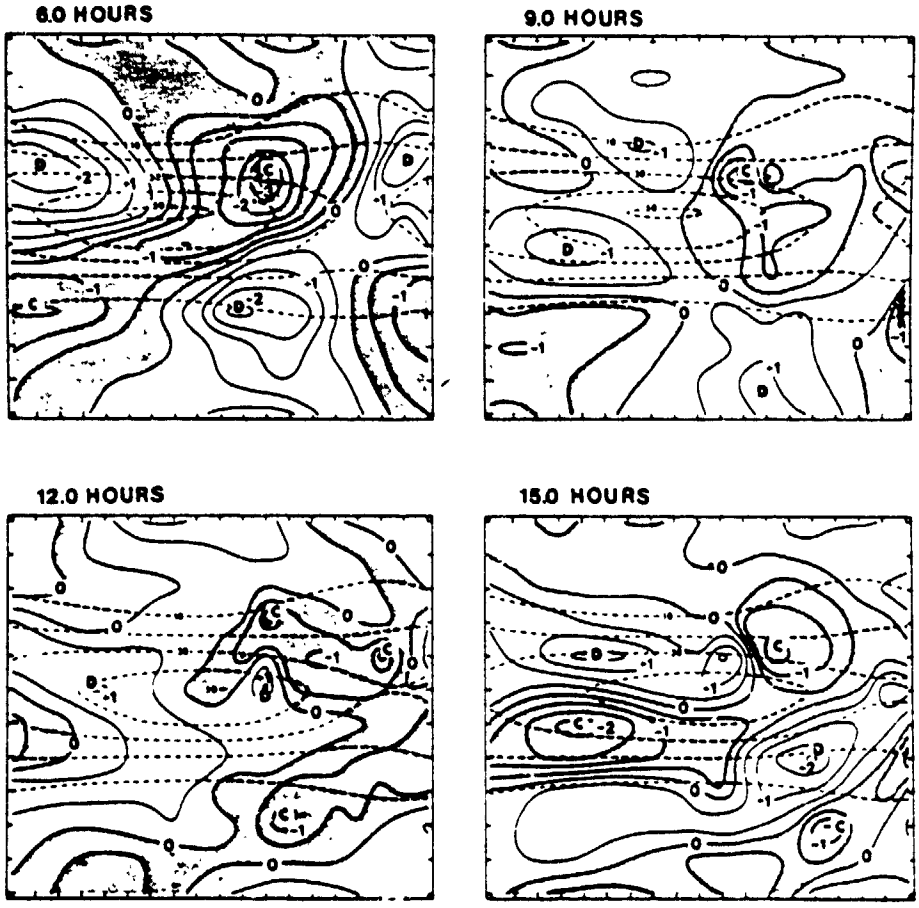


FIG. 6. Isotachs (dashed lines, m s^{-1}) on 340 K isentropic surface and $\dot{\sigma}$ at interface (solid lines, 10^{-6} s^{-1}). Shaded region represents negative $\dot{\sigma}$ or convergence (C), unshaded region positive $\dot{\sigma}$ or divergence (D). A multiplication of this field by $g^{-1}\rho^{*}$ yields the vertical mass flux through the interface.

(Fig. 6) shows a large-scale, four-cell pattern representative of propagating jet streaks (Bjerknes, 1951; Riehl *et al.*, 1952) and a transient feature stemming from the geostrophic initialization. The two centers of rising motion are generally associated with the cyclonic vortices while the two centers of sinking motion are generally associated with the anticyclonic vortices. The development of $\dot{\sigma}_B$ represents, in part, the vertical branches of the transverse circulations which produce a significant exchange of mass and momentum between the two model domains. The transient feature of the $\dot{\sigma}_B$ field is associated with the boundary layer convergence zone which lies along the anticyclonic side of the streak at 6.0 h. During the following 9 h period this feature propagates rapidly southeastward at an average rate of 50 m s^{-1}. By 15.0 h, the original convergence band has moved far ahead of the streak and a four-cell divergence pattern has redeveloped beneath the streak itself. The convergent band that rapidly propagated to the southeast appears to be a gravity inertia wave which stems from the dynamic imbalance between the mass and momentum distributions intitialized by the geostrophic approximation (Charney, 1955; Phillips, 1960). Both the large-scale and transient features of the divergence fields evident in this experiment were similar to the patterns in Gall's (1972) experiments. Since Gall's (1972) experiments were conducted with an inactive boundary layer and no vertical exchange of mass through the interface, the origin of the gravity inertial wave is in all likelihood due to geostrophic initialization used in both experiments and not due to the intitialization of emerging grid points in this experiment.

The results from model Experiment E2, designed to test the hybrid model with diffusion, are basically the same. With the addition of diffusion, the maximum wind velocity and propagation rate for the jet streak decreased more rapidly than for the nondiffusion Experiment E1. The maximum velocity decreased from 54.8 to 50 m s^{-1} by 12 h, and to 43 m s^{-1} by 30 h, as the average propagation rate decreased from 30 to 18 m s^{-1} to 12 m s^{-1} for the first, second and third 12 h periods, respectively. This change in the character of the jet streak, however, was not reflected in the basic development of the surface pressure cells nor in the vertical motion through the interface, both of which were similar to the results from Experiment E1.

b. Conservation of mass and momentum

An important aspect of the initial hybrid model experiments is to demonstrate that physical properties are conserved with respect to transport processes for the entire model domain. Mass conservation for the total as well as for the isentropic domain can be checked by calculating the area-weighted average pressure on the interface $\bar{p}_B$. Given hydrostatic balance and the lateral boundary conditions specified in Section 3h, $\bar{p}_B$ and $\bar{p}_s$ must remain constant if mass is conserved. The deviations of $\bar{p}_B$ and $\bar{p}_s$ from their initial values of 800 and 1000 mb were less than 3×10^{-3} mb during the entire 36 h period, the magnitude that is expected from round-off error. This respresents less than a 4×10^{-4}% change of $\bar{p}_B$ and $\bar{p}_s$ in spite of significant mass transport through the interface (Fig. 6).

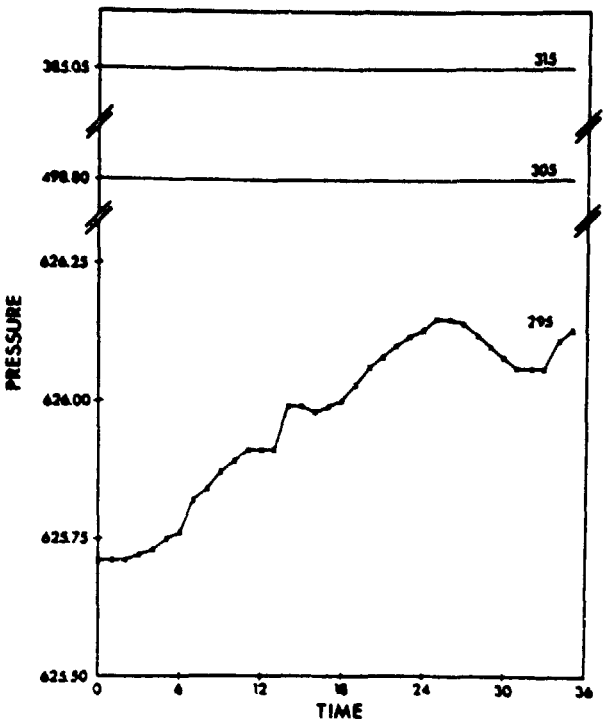


FIG. 7. Average pressure $\bar{p}_k$(mb) for three isentropic surfaces. The 315 and 305 K surfaces remain above the boundary layer while 295 K intersects the boundary layer.

Since isentropic, layer-averaged, hydrostatic mass distribution is conserved under adiabatic conditions, the area-averaged pressure $\bar{p}_k$ for each isentropic surface should remain constant whether it remains above the interface or intersects the sigma-coordinate boundary layer. The area-average pressure remained constant for all isentropic surfaces completely above the interface but increased in these experiments for isentropic surfaces which intersected the interface, e.g., the mean pressure on the 295 K surface increased by 0.5 mb over 36 h (Fig. 7). The general increase of $\bar{p}$ for 295 K is related in part to the interpolation of θ to the interface and the subsequent interpolation of p to isentropic surfaces within the boundary layer and truncated domain. This problem is accentuated during the emergence of grid points into the isentropic domain from the sigma boundary layer. Note in the results (Fig. 7) that jumps of $\bar{p}_k$ occurred at 7, 15 and 35 h in conjunction with the emergence of a grid point on the 300 K surface. Another factor in the overall change of $\bar{p}_k$ for isentropic surfaces which intersect the sigma

domain is the truncation error involved with predicting θ in the sigma domain (II4). A finite form of (II4) does not maintain constant mass between isentropic surfaces within any model expressed in sigma coordinates. Thus, while mass is conserved for the entire model domain, the proper vertical distribution of mass with respect to the isentropic surfaces within the sigma domain is not exactly maintained even for adiabatic flow.

Momentum budgets were utilized in the initial model experiments to insure that the vertical transports between the two model domains were matched exactly and to check that horizontal transports between truncated grid volumes were also matched. A comparison of the u and v momentum tendencies with its forcing terms indicated a perfect match for all model regions. The changes in the u momentum were smooth and fluctuated with periods ranging from 5 to 12 h, while the v momentum tendencies were more varied. Calculations of momentum and energy budgets were made for the sigma domain, the "truncated" isentropic domain immediately above the interface and the "free" isentropic domain in which the grid volumes remain entirely above the interface. Inspection of the contribution of the individual forcing terms in each model region indicated that the net momentum changes in each region were dominated by the generation due to the ageostrophic wind component. Further details are discussed by Uccellini *et al.* (1977).

c. *Energy budget, Experiment E1*

As an independent check of the model performance, the kinetic energy equations for the isentropic and sigma domains are evaluated by

$$\frac{\partial}{\partial t}(\rho J_\sigma J_\sigma k) = -\frac{1}{a}\left[\frac{\partial}{\partial \Lambda}(a^2 \rho J_\sigma k u) + \frac{\partial}{\partial \Phi}(a \rho J_\sigma J_\sigma k v)\right] - \rho J_\sigma J_\sigma k \dot{\sigma}\Big|^{\sigma_s} - \rho J_\sigma J_\sigma \frac{1}{a}\left[\left(u J_\sigma^{-1}\frac{\partial \psi}{\partial \Lambda} + v\frac{\partial \psi}{\partial \Phi}\right)\right] + \nu(u\nabla^2 u + v\nabla^2 v), \quad (18)$$

$$\frac{\partial}{\partial t}(\rho J_\sigma J_\sigma k) = -\frac{1}{a}\left[\frac{\partial}{\partial \Lambda}(a^2 \rho J_\sigma k u) + \frac{\partial}{\partial \Phi}(a \rho J_\sigma J_\sigma k v)\right] - \frac{\partial}{\partial \sigma}\rho J_\sigma J_\sigma \dot{\sigma} - \rho J_\sigma J_\sigma \frac{1}{a}\left[\theta\left(J_\sigma^{-1} u \frac{\partial \pi}{\partial \Lambda} + v\frac{\partial \pi}{\partial \Phi}\right) + \xi\left(u J_\sigma^{-1}\frac{\partial z}{\partial \Lambda} + v\frac{\partial z}{\partial \Phi}\right)\right] + \nu(u\nabla^2 u + v\nabla^2 v), \quad (19)$$

where $k = \frac{1}{2}(u^2+v^2)$. The vertical flux in (18) only applies to truncated grid volumes adjacent to the interface. After integration over each hour, both sides are compared to check if the tendencies match the forcing terms. For Experiment E1, net changes of the integrated kinetic energy (K) are due to conversion from the available potential energy (A) by ageostrophic winds, and by vertical flux through the interface in the sigma and truncated regions.

The available potential energy (Lorenz, 1955; Johnson, 1970)

$$A = \left[\frac{c_p}{g p_{00}^{\kappa}(1+\kappa)}\right]\int_0^{\infty}\int_A (p^{1+\kappa} - \bar{p}^{A\,1+\kappa})\,dA\,d\theta, \quad (20)$$

is calculated every hour, where $\bar{p}^A$ is the area-averaged pressure for each layer. For adiabatic experiments, the net hourly A changes should match the K hourly changes exactly. The method of calculating the available potential energy is discussed by Uccellini *et al.* (1977).

The comparison of the kinetic energy tendencies with the forcing terms indicates a near match for the separate model regions (Fig. 8). The kinetic energy tendencies reveal an oscillation in the free isentropic region with a period near 12 h in response to the geostrophic initial conditions (Charney, 1955). The net kinetic energy decreases in the free and truncated regions and increases a lesser amount in the sigma domain, resulting in a net decrease of K for the entire model domain. The decrease of K in the upper troposphere is reflected by the decrease in magnitude of the jet streak, while the K increase in the sigma domain is coincident with the development of the vortices. The kinetic energy changes in all three regions are dominated by the conversion terms as the net vertical transport through the interface is nearly zero (Fig. 9). For the sigma domain the conversion is related to $\mathbf{V}\cdot\nabla_\sigma \pi$ which remains positive for the entire period and offsets the smaller negative tendencies related to the $\mathbf{V}\cdot\nabla_\sigma z$ term.

The available potential and kinetic energy tendencies, compared in Fig. 10A, are in close agreement indicating the net decrease of kinetic energy is being accounted for by the conversion to available potential energy. Thus the net K changes are related to physical processes and not to numerical deficiencies. The small differences which exist result in a net gain of the total energy for the first 30 h and a small net loss for the last 6 h (Fig. 10B). The causes for the differences may be related to the lack of conservation of area-averaged pressures for isentropic surfaces which intersect the boundary layer (Section 4b). Conservation of the area-average pressure under adiabatic conditions is a basic factor in the theory of available potential energy. The lack of conservation in part stems from the interpolation of information from sigma levels to isentropic surfaces in the boundary layer and the extrapolation of θ to the earth's surface

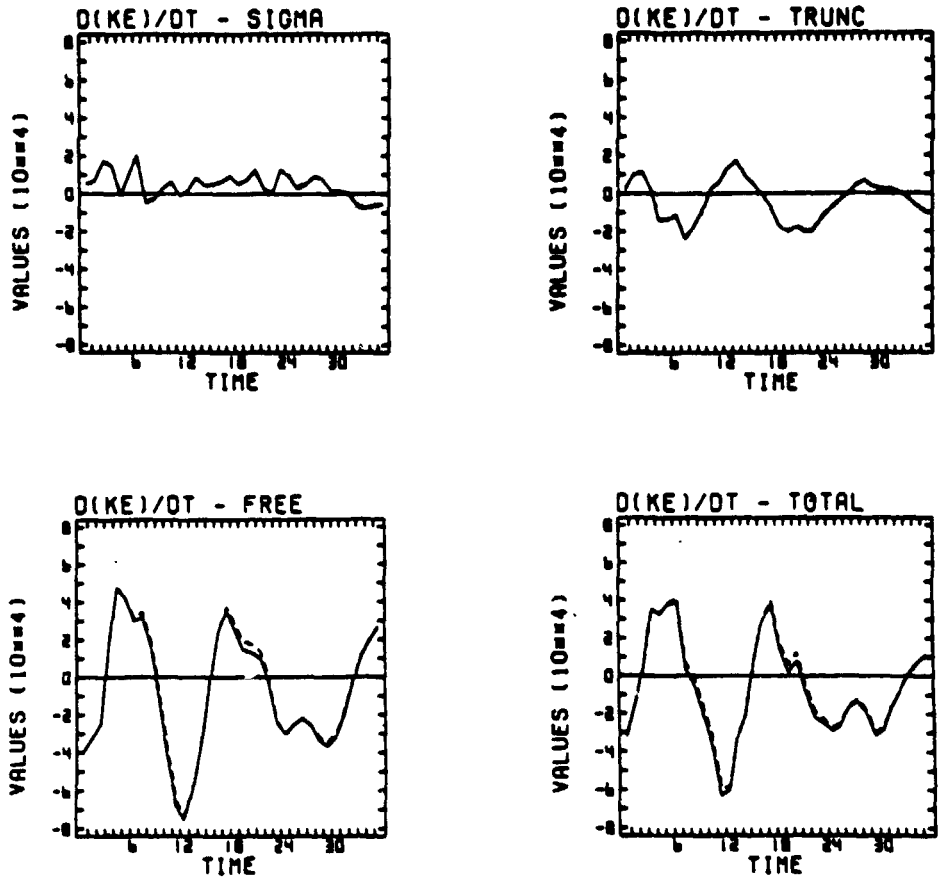


FIG. 8. Diagnostic comparisons of kinetic energy tendency per unit area (g s^{-3}) for Experiment E1. Left side of kinetic energy equation (solid) and right side (dashed) are shown for three model regions and total model domain.

needed to calculate the available potential energy (Uccellini *et al.*, 1977). It may also stem from implicit truncation of the pressure fields and potential temperature within the sigma domain. In any event, the deviation of the total energy (Fig. 10B) is three orders of magnitude less than the initial total energy and is considered to be minor for integrations of limited durations. In Experiment E2, which included diffusion, the kinetic energy tendency was an order of magnitude larger than in Experiment E1 and was predominantly negative. Significant changes in the tendency occurred in the "free" and "truncated" isentropic regions where the tendencies fluctuated with the same period as before, but remained negative throughout. The kinetic energy tendency in the sigma domain was the same as in Experiment E1 but was relatively less important, being a full order magnitude less than the tendencies in the isentropic domain (Uccellini *et al.*, 1977).

d. Model performance at the interface

Perhaps the most sensitive region of the hybrid model is the interface and the truncated isentropic grid volumes immediately above it. Although predictive equations are not evaluated at the interface, the vertical fluxes of mass, momentum and energy through the interface serve as the upper and lower boundary conditions for the sigma and isentropic domains, respectively. These fluxes, common to the two model domains, must match exactly to conserve the physical properties. The vertical fluxes of momentum and kinetic energy through the interface are matched, as a net gain or loss in one domain is exactly offset by a net loss or gain in the other domain. The fact that $\bar{p}_s^{\,4}$ also remained constant (Section 4b) verifies that the vertical mass fluxes calculated at the interface do not generate spurious sources or sinks of mass. Thus, conservation of the physical properties, with respect to vertical transport through the interface, is assured for this hybrid model approach.

In other hybrid models, adjustments of predicted variables were needed to maintain a smooth transition between model domains. Such adjustments may lead to spurious gravity wave modes and resulting pressure oscillations (Friend, 1976). In this model approach, the

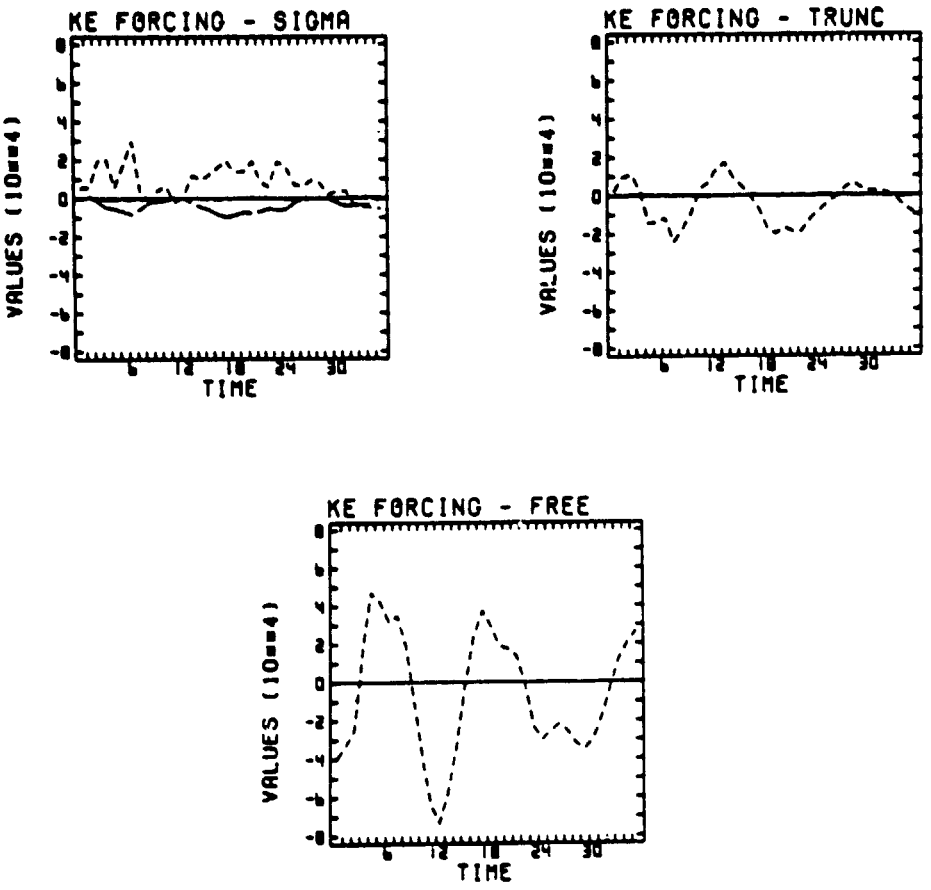


FIG. 9. The "forcing" terms of kinetic energy tendency per unit area (g s^{-3}) for Experiment E1. In sigma domain, forcing due to the $V \cdot \nabla_\sigma v$ term is shown by short dashes, forcing due to $V \cdot \nabla_\sigma z$ term by the long dashes. For truncated and free isentropic regions, ageostrophic forcing is represented by the short dashes.

predictive equations are evaluated on either side of the interface, with new values of θ, u and v interpolated to the interface level. This interpolation procedure is designed to maintain a smooth transition between the two model domains without adjusting the solutions for predicted variables. Fig. 11 portrays a sequence of cross

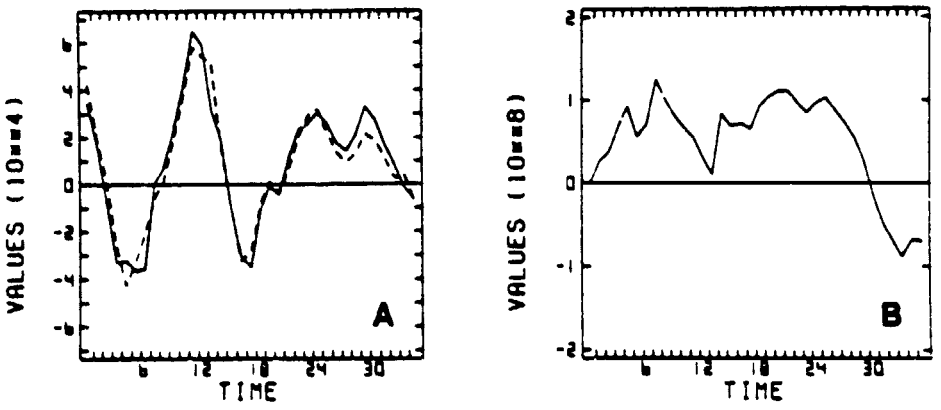


FIG. 10. (A) Available potential energy per unit area tendency (dashed) and negative kinetic energy per unit area tendency (solid) for total model domain (g s^{-3}) for Experiment E1. (B) Deviation of total energy per unit area $(A+K)$ from initial value $(A+K)_{t=0}$ (g s^{-3}) for Experiment E1.

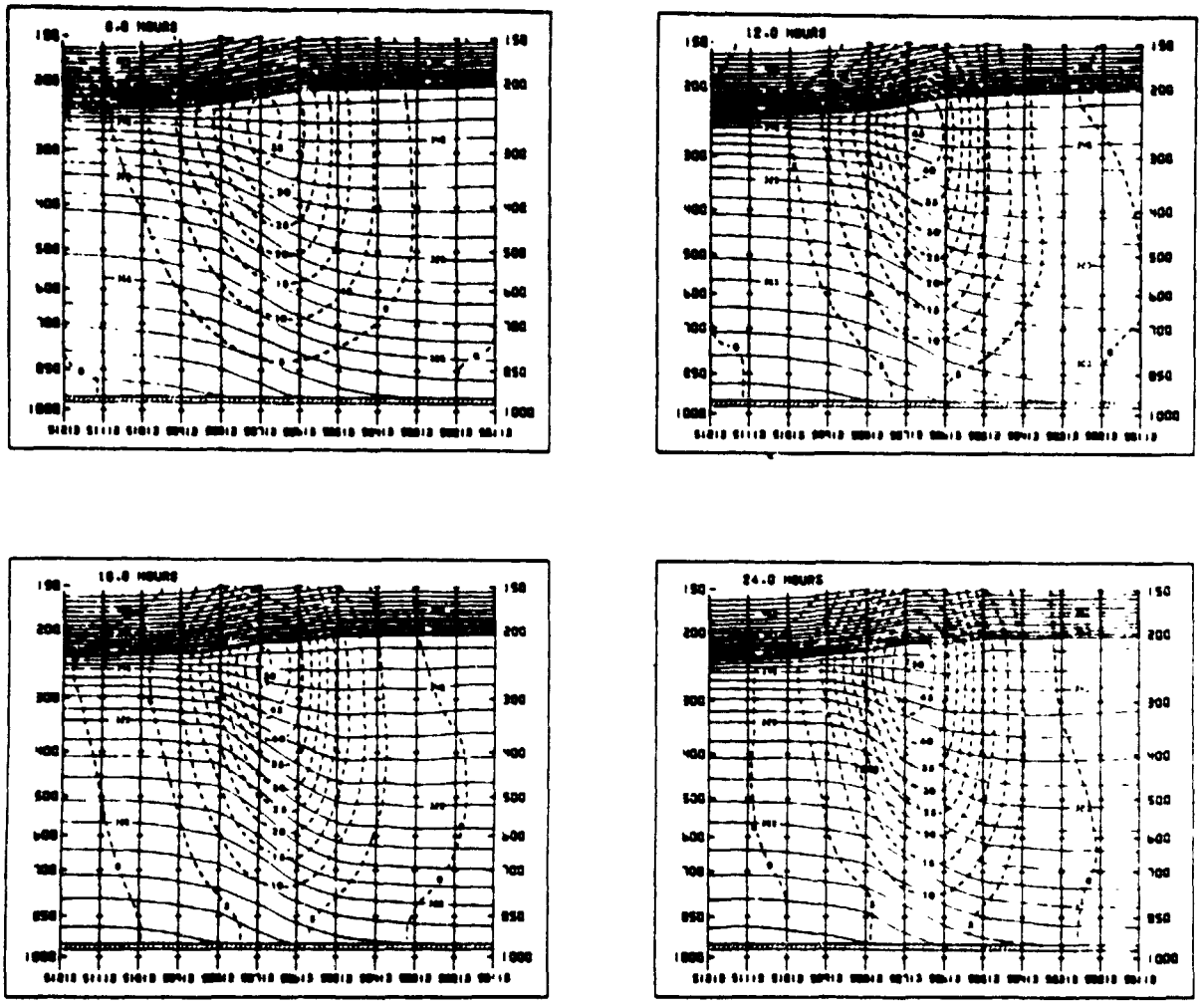


FIG. 11. Sequence of cross sections along column 13 for times 6, 12, 18 and 24 h. Isentropes, solid lines, isotachs (m s^{-1}), dashed lines, direction of flow is into page.

sections (6 h apart) in the exit region of the jet streak and illustrates the success of the interface interpolation procedure in maintaining a smooth transition of the mass and wind fields between the two model domains. No discontinuities develop either in the potential temperature or in the isotach fields which intersect the interface (~800 mb), even in regions where significant vertical fluxes exist.

e. Diagnostics on the initialization of grid points emerging into the isentropic domain

Four grid points were isolated in both model experiments to test and illustrate the stability of the predictive equations in different regions of the model. Grid point 1 (300 K surface) is located in the truncated isentropic region and remains above the interface for the entire experiment. Grid point 2 (290 K surface) is also located in the truncated isentropic region and lies immediately below grid point 1. It emerges twice and remains above the interface after the second emergence. Grid point 3 (0.25 σ surface) is located in the sigma domain immediately below the interface and below the position of grid points 1 and 2. These three grid points were chosen to illustrate the effects of the initialization procedure for emerging grid points (Appendix D) on the different model regions near the interface. Grid point 4 (340 K surface) is located in the free isentropic region at the jet streak level and was chosen primarily for a comparison of the momentum tendencies and wind components between the different model domains.

Figs. 12 and 13 illustrate the momentum tendencies and wind components for the four grid points and the mass Δp for grid points 1 and 2 for Experiment E1. In each figure the profiles of the momentum tendencies

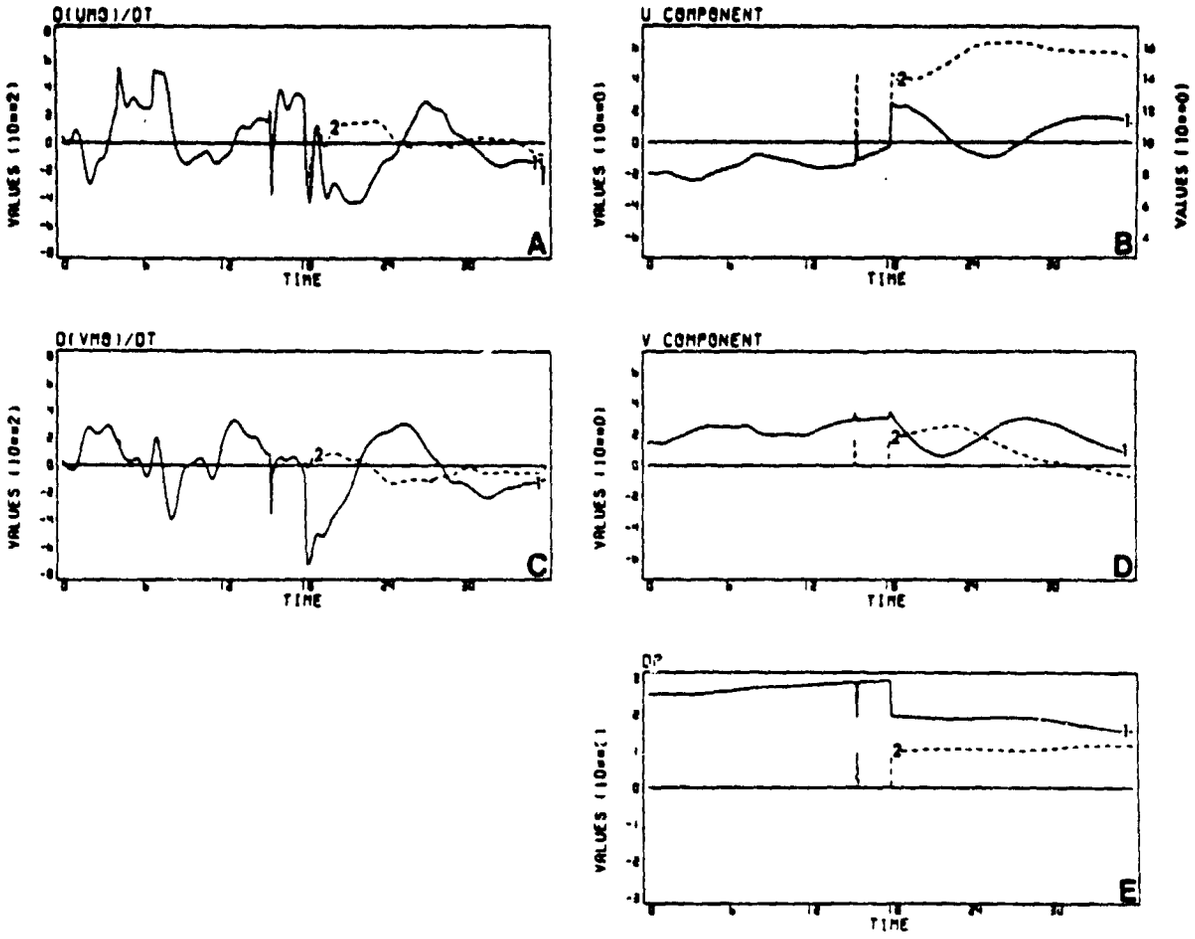


FIG. 12. Grid-point diagnostics for grid points 1 (grid point immediately above the interface preceding emergence of grid point 2) and 2 (emerging grid point immediately below 1), Experiment E1. A and C. u and v momentum tendencies per unit area (g s^{-2} m^{-1}); B and D. u and v wind components (m s^{-1}); E. pressure increment Δp (mb). In B, the u component for grid point 1 is scaled on the right, for grid point 2 on the left.

are to the left and those of the wind components and mass are to the right. The u component for grid point 1 (solid lines, Fig. 12B) is scaled on the right border and the u component for grid point 2 (dashed line, Fig. 12B) on the left border. This separate legend was used to preserve equal scaling over a different range for both wind components and to compare the variability of u and v among the grid points after the emergence of grid point 2.

Fig. 12 reveals the redistribution of mass during the emergence of grid point 2 into the isentropic domain that is necessary for mass conservation (Appendix D). The adjustments of the wind components which insure momentum conservation produce an increase for both u and v at grid point 1 followed by a noticeable oscillation (Figs. 12B and 12D). The wind changes force an imbalance between the wind components and the pressure gradient force, which is evident by the sharp fluctuations in these momentum tendencies at 18 h in E1 (Figs. 12A, 12C). The amplitude of the inertial wind oscillation occurring for grid point 1 is 1.5 m s^{-1}. With diffusion included, the amplitude of the fluctuations in the momentum tendencies and the wind components which result from the emergence of grid point 2 is reduced by nearly a third (Uccellini *et al.*, 1977).

The emergence of grid point 2 is also responsible for a sharp decrease in the u momentum tendency for grid point 3, located in the sigma domain (Fig. 13A). The oscillation in the u component reaches amplitudes of 2 m s^{-1}. A close inspection of the interface pressure tendencies below the emerging grid point and for the four surrounding grid points (Fig. 14) reveals that the changes in the u momentum are a result of a rapid increase in $\partial\pi/\partial x$, a part of the east–west pressure gradient force in sigma coordinates. The pressure at the grid point to the east increases by 0.75 mb in 15 min, while to the west it decreases by 1.00 mb in 20 min.

The cause of the sudden change in $\partial\pi/\partial x$ after the emergence of grid point 2 is related to the method by which horizontal transports are calculated in the trun-

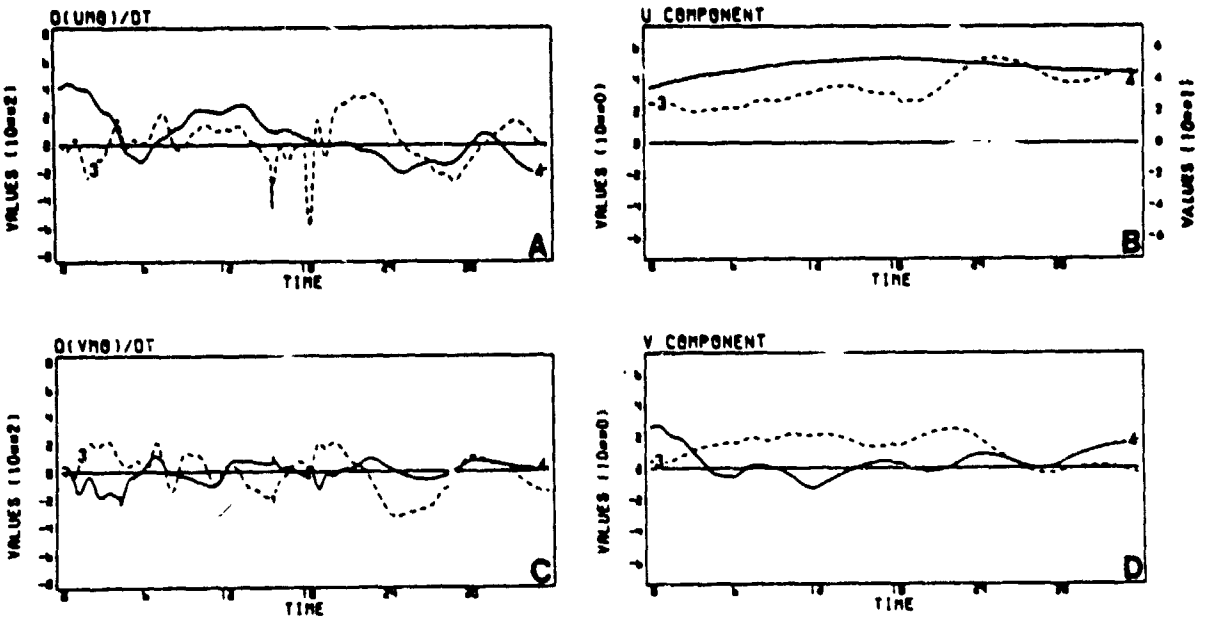


FIG. 13. Grid-point diagnostics for grid points 3 (in the sigma domain immediately below 1 and 2) and 4 (in the upper isentropic domain), Experiment E1. A and C: u and v momentum tendencies per unit area (g s^{-2} m^{-1}); B and D: u and v wind components (m s^{-1}). In B, u component for grid point 4 is scaled on the right and grid point 3 on the left.

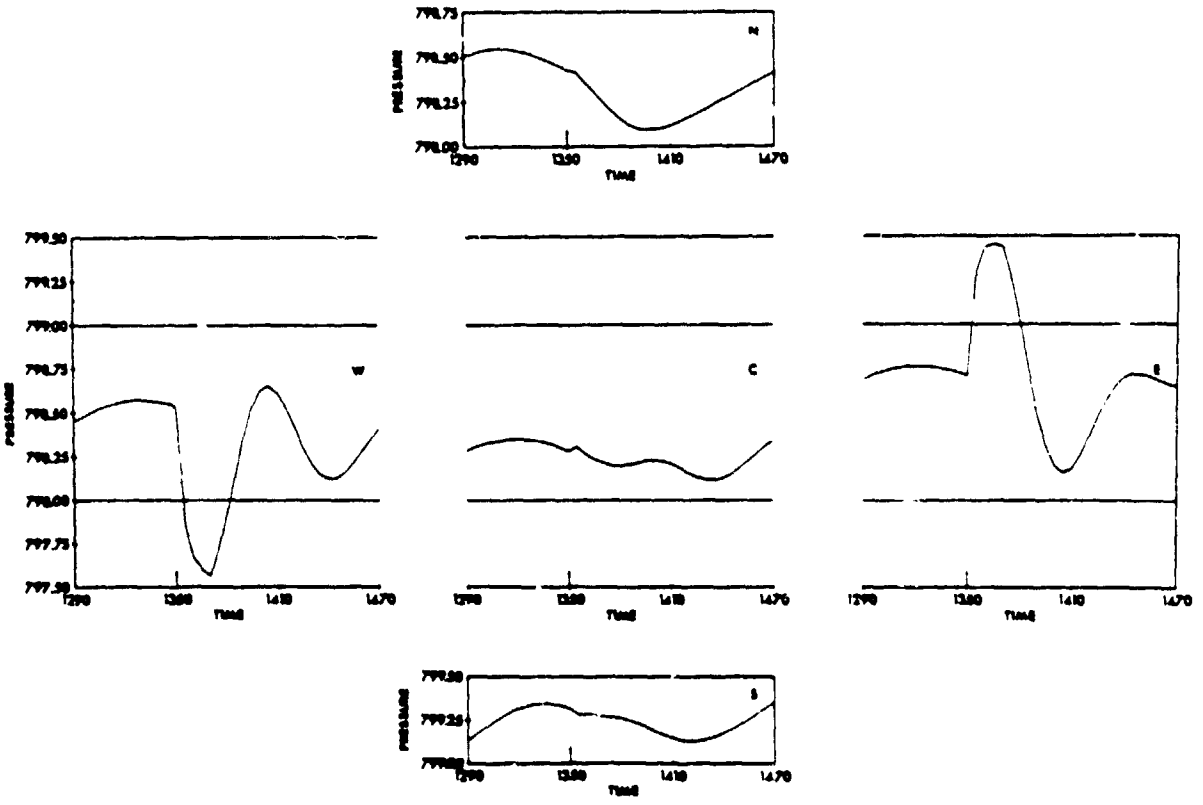


FIG. 14. Interface pressure tendencies for grid point (C), immediately beneath the grid point (2) emerging into the isentropic domain and for the four grid points immediately surrounding C. Pressure (mb), time (min), arrow indicates time of emergence of grid point 2.

cated isentropic region. The addition of grid point 2 changes the method of calculating transport at the center column from a "normal" truncated box to an "A2" truncated box (see Appendix B). This change involves calculating transports in two grid volumes rather than the one truncated grid volume prior to emergence and utilizing the increased wind components at grid point 1 which result from the adjustment procedure designed to conserve momentum (Appendix D). The sudden change in the method for calculating transport at the grid column after the emergence of grid point 2 combined with the increase of the u component at grid point 1 (Fig. 12B) increases the west to east mass transport and thus forces the spurious pressure gradient within the sigma domain. The vertical extent of the oscillation is limited to the layers adjacent to the new grid element. The upper tropospheric grid point (grid point 4, Fig. 13) shows no evidence of oscillation after emergence of grid points near the interface. Although the magnitude of the oscillation is small, decreases rapidly after one cycle in this experiment and is significantly reduced with diffusion included, cases where large wind components and significant vertical shears exist could pose problems. The implementation of a time-weighted transition between the two methods of calculating transport for emerging or submerging grid points will be tested in future experiments.

5. Summary and future research

Hybrid models, which combine an isentropic representation of the free atmosphere with a sigma-coordinate boundary layer, are designed to take advantage of the strengths of the different coordinate systems. The packing of isentropes in baroclinic zones increases the vertical resolution in regions where large vertical wind shears exist and where middle and upper tropospheric dynamical processes tend to be amplified. Yet, the intersection of isentropes with the ground, the existence of superadiabatic and adiabatic conditions, and the general lack of vertical resolution near the earth's surface in less stable regions inhibit incorporating boundary layer dynamic and thermodynamic processes into isentropic models. Without an adequate representation of the boundary layer, isentropic models are limited in their application to subsynoptic- and synoptic-scale diagnostics and forecasts. These deficiencies of isentropic coordinates are avoided by utilizing sigma coordinates within the boundary layer which maintain a uniform grid spacing throughout, avoid the intersection between coordinate surfaces and the ground, and thus provide a suitable framework for incorporating boundary-layer processes into the model.

The primary problem with hybrid models is matching the boundary conditions across the interface while providing for full interaction between the model domains without introducing extraneous wave modes through spurious sources of mass, momentum or energy. The solution presented in this paper matches boundary conditions at the interface using the flux form of the primitive equations. The model design described in this paper includes a vertically staggered grid, which provides for easy implementation of the matching boundary conditions by direct calculation of vertical transport through the interface without need for parameterization. The conservation of physical properties associated with transport processes is therefore assured, even in the region of truncated grid volumes near the interface. This grid structure also avoids the integration of predictive equations on the interface since the wind and temperature distributions are determined by interpolation between the two model domains. This eliminates the possibility of predictive solutions at grid points on the interface diverging from grid-point solutions immediately above, as well as the need for artificial adjustments to maintain a smooth transition between the isentropic and sigma domains.

The initial model experiments, designed to test the feasibility of the hybrid model approach and to study the model behavior near the interface, were applied to a jet streak propagating in a zonal channel. The results of these experiments with and without diffusion show the following:

1) The model appears stable up to the time of termination without spatial filtering or diffusion.

2) The model integrations of the propagating jet yield nearly identical results to Gall's (1972) model experiments which did not include an active sigma boundary layer and correctly account for the mass and momentum adjustments associated with a jet streak of finite length.

3) No spurious sources or sinks of mass occur within the *entire* model domain, although the correct vertical distribution of mass within the sigma domain is not exactly maintained. The significant changes in momentum and energy are accounted for by the individual physical forcing functions, while minimal truncation errors stem from the initialization of emerging grid points.

4) The vertical transport of mass, momentum and energy through the interface from one domain is exactly matched by transport into the other domain, thus insuring the conservation of basic physical properties with respect to transport processes.

5) A smooth transition in the mass and wind fields can be maintained across the interface without need for artificial adjustments.

6) The initialization of emerging grid points in the truncated isentropic region, based on a redistribution of mass and momentum that satisfies conservation principles, results in pressure and wind perturbations in the truncated grid volumes and the sigma grid volumes below. These perturbations associated with the sudden imbalance between the wind and mass fields are due to the change in methods of calculating transports in the

truncated region that result from having different types of grid elements before and after grid point emergence. The nonconservative adjustments of the energy distribution ψ also contribute to an artificial imbalance between the mass and wind fields. Although the perturbations are stable and quickly damp, the need to initialize grid points in the sensitive model region above the boundary layer is a problem with hybrid models which merits further attention.

In general, the three-dimensional hybrid model approach is feasible and can be applied to study a variety of atmospheric phenomena.

Future model experiments should include testing alternative techniques for the initialization of emerging grid points and the calculation of the pressure gradient force for partially submerged grid volumes in an effort to reduce the magnitude of spurious pressure and wind perturbations. While the flux form of the equations utilized in this model insures conservation with respect to transport processes (i.e., convective flux), the finite-difference techniques for the computation of the pressure and viscous forces do not insure conservation with respect to the nonconvective flux of momentum and energy at the interface (Johnson, 1977). Finite-difference forms for the pressure and viscous forces that insure conservation with respect to nonconvective flux across the interface as well as for truncated grid volumes should be tested. Terrain will also be added in an effort to ascertain the strengths and weaknesses of the hybrid model approach for mountainous regions. Moist thermodynamics will be added and the boundary layer should be improved by adding sensible heating and a realistic parameterization of friction. The improvements will be important for future studies of the three-dimensional mass and momentum adjustments associated with propagating jet streaks, frontogenesis and cyclone evolution.

Acknowledgments. The authors wish to acknowledge the useful discussions with Drs. Rainer Bleck, Robert E. Gall and Dennis Deaven during the course of this work which have contributed to our research efforts, and the reviewers who made many useful suggestions which helped clarify this manuscript. We express our appreciation to Mr. Thomas Whittaker for his computing assistance, Mr. John Stremikis for his assistance in drafting the figures and editing the final manuscript and Ms. Jean Johnson for typing the final manuscript. This research was sponsored by the Atmospheric Sciences Section, National Science Foundation, under Grants ATM75-23223 and ATM77-22976.

APPENDIX A

List of Symbols

A. *General*

(f)	general physical property (e.g., mass, momentum, energy)
$\mathbf{U}$	two-dimensional horizontal wind vector
u	wind component in x direction
v	wind component in y direction
f	Coriolis parameter
p	pressure
π	Exner function $[=c_p(p/p_{00})^{\kappa}]$
a	radius of earth, also incremental distance in Appendix B
ϕ	latitude
T	temperature
ρ	density
θ	potential temperature
$\partial\Lambda$	latitude grid increment
$\partial\Phi$	longitude grid increment
J_*	Jacobian transformation to modified spherical map $[=(a^2\cos\phi/\cos 45°)]$
Q	latent heat
F_u, F_v	generalized friction
h	grid length
K	integrated kinetic energy; also proportionally constant for eddy diffusivity
A	available potential energy
D	deformation
ν	viscosity coefficient
$\oint dL$	line integral
$\int_A dA$	area integral
$\mathbf{n}$	unit normal vector

B. *Isentropic domain*

ρJ_θ	Jacobian transformation $[=-1/g(\partial p/\partial\theta)]$
Δp	pressure difference between isentropic surfaces
ψ	Montgomery streamfunction
$\dot{\theta}$	$d\theta/dt$
θ_u, θ_L	upper and lower limits of vertical integration
$\bar{\theta}$	mean potential temperature for truncated layers

C. *Sigma domain*

ρJ_σ	Jacobian transformation $[=1/g(\partial p/\partial\sigma)]$
p^*	gJ_σ, set equal to 200 mb
z	height
σ	normalized pressure coordinate in boundary layer
$\dot{\sigma}$	$d\sigma/dt$

D. *Subscripts*

T	top of model
s	earth's surface
B	interface

η	generalized coordinate
0	central grid point
1, 2, 3, 4	east, north, west, south grid points
i, j, k	row, column, vertical indexes

E. Initialization

Y	width of model domain
X	length of model domain
y_0	center of channel
y_s	special scaling factor to enhance baroclinic structure in the middle of the channel
γ	lapse rate $[=-\partial T/\partial z]$
γ_d	dry lapse rate $[=g/c_p]$
$\bar{\gamma}$	mean initial lapse rate at various levels
$\Delta\gamma$	amplitude of wave for initial lapse rates
$\bar{\theta}_s$	mean surface potential temperature
$\Delta\theta_s$	amplitude of wave for initial surface potential temperature
A	coefficient which influence initial north–south potential temperature gradient
B	coefficient which determines amplitude of initial wave

F. Constants

c_p	specific heat capacity of dry air at constant pressure
p_{00}	reference pressure (1000 mb)
R_d	gas constant for dry air
κ	R_d/c_p

APPENDIX B

Application of Box Method to Truncated Grid Volumes

Fig. 15 illustrates several examples of truncated boxes located around grid point $x_{(1,0)}$. The truncated box extends from the interface to the upper "solid" surface in all examples. The flux through the left-hand side of the box is determined by a modification of the averaging technique described above. The fluxes are calculated from the products of Δp between the interface and the upper "solid" surface, and a linearly interpolated value of u, v and the property (f) to a mean isentropic level $(\bar{\theta})$ for each grid column. Note that in this determination of Δp, the lower solid isentropic surface immediately above the interface has no role in these calculations and is essentially ignored. The flux through the wall is then determined using the averaging technique described earlier. The same procedure is applied to the right side of the grid volume where either the bottom "solid" surface or both the bottom and the "dashed" surface intersect σ_B (cases A1 and A2; Fig. 15).

When the top of the grid volume centered at $x_{(1,0)}$ (case A3; Fig. 15) also intersects the interface within the $2\Delta x$ region, the flux through the wall is determined by a linear extrapolation of the flux which exists at the center grid column. In the example in Fig. 15, the flux through the right wall is given by

$$[(\Delta pu)(f)]_{\text{right wall}}=\frac{a}{[DX]}[(\Delta pu)(f)]_{\text{center column}}, \quad (21)$$

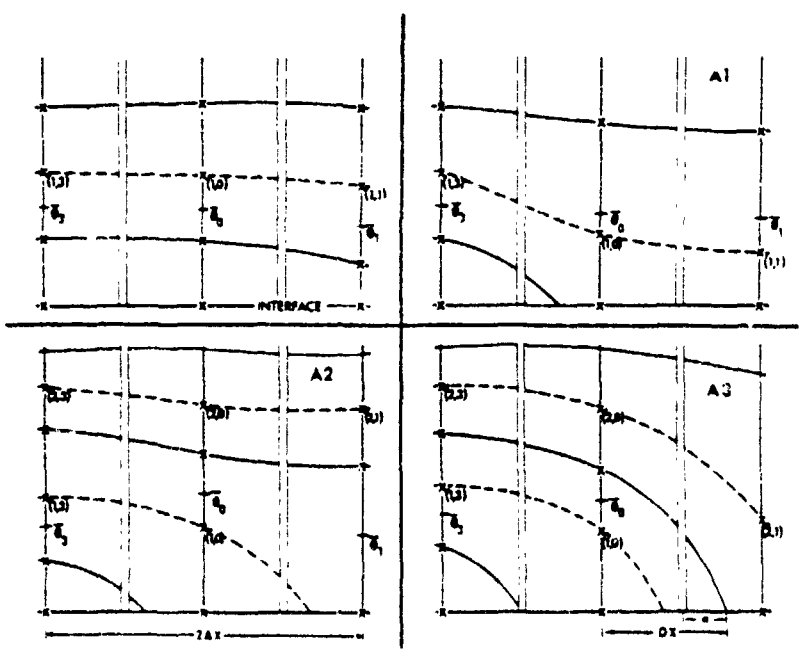


FIG. 15. Examples of "truncated" isentropic boxes that require modification of finite differencing. Top left shows an example of truncated box located at $x_{(1,0)}$ with σ_B as lower boundary. A_1, A_2 and A_3 are examples of truncated boxes where isentropic surfaces within an isentropic grid volume centered at $x_{(1,0)}$ intersect the lower boundary σ_B.

where [DX] is the distance from the grid point on the interface directly below the central grid point of the box to the point where the top surface and the interface intersect. The incremental distance a, the length between the grid wall and the point of intersection, is given by

$$a=([\mathrm{DX}]-\tfrac{1}{2}\Delta x). \tag{22}$$

If $a\leqslant 0$, the top surface intersects within the box itself and the flux through the right side is set equal to zero.

For cases A2 and A3 the transport through the right wall has to be accounted for in the adjacent box located about $x_{(2,1)}$ (Fig. 15) to insure the conservative nature of the method. This is accomplished by specifying the transport through the left wall of the adjacent box $x_{(2,1)}$ as the sum of the individual transports calculated to the right of $x_{(2,0)}$ and $x_{(1,0)}$. This does not imply the existence of mass transport through isentropic surfaces but only horizontal transport from two grid volumes to an enlarged, adjacent grid volume for which the mass and momentum tendencies are represented at grid point $x_{(2,1)}$.

APPENDIX C

Initialization of Model

The basic structure for the initialization of the hybrid numerical model is identical to Gall's (1972) model for a jet streak. However, in this application, provisions had to be developed to extend the baroclinic structure throughout both domains of the model. The initial structure for the atmosphere was specified by the surface distribution of potential temperature

$$\theta_s=\theta_{s_0}+\Delta\theta_s\tanh\left\{\frac{y-y_0}{y_*}\left[A+B\sin\left(\frac{2\pi x}{X}\right)\right]\right\} \tag{23}$$

and the lapse rate distribution

$$\gamma=\gamma_0+\Delta\gamma\tanh\left\{\frac{y-y_0}{y_*}\left[A+B\sin\left(\frac{2\pi x}{X}\right)\right]\right\}, \tag{24}$$

where the subscript 0 indicates values at the center of channel, and the amplitudes are specified by $\Delta\theta_s$, $\Delta\gamma$, A and B (see Table III). A scaling parameter y_* compacts the baroclinic structure to the center of the channel.

In this model the surface potential temperature is applied to the lower boundary $\sigma=1$, while the vertical variation of the potential temperature within the sigma domain is specified by vertically integrating

$$\frac{\partial(\ln\theta)}{\partial\sigma}=\frac{\kappa p^*}{p}\left(\frac{\gamma-\gamma_d}{\gamma_d}\right). \tag{25}$$

A uniform pressure distribution of 1000 mb at the earth's surface was assumed for the initial structure. The height z is initialized in the sigma domain from the hydrostatic relation (II6). Although in these preliminary experiments the height of the earth's surface was assumed to be zero, topography could be included without significant modification.

In the isentropic domain, the pressure was determined by vertically integrating

$$\frac{\partial(\ln p)}{\partial\theta}=\frac{-g}{R_d\theta(\gamma_d-\gamma)}, \tag{26}$$

up to isentropic level 3 (see Fig. 1) where the lapse rate γ is specified by (24). The layer between levels 2 and 3 was assumed to be a transition region between the troposphere and stratosphere (see Fig. 1). The pressure distribution within this layer and above was prescribed by alternate methods to insure a realistic decrease of wind speed with elevation in the stratosphere, thereby restricting the jet maximum to the tropopause region (Uccellini *et al.*, 1977).

With the mass distribution specified in both model domains, the winds were initialized by the geostrophic approximations, which are then utilized to initialize $\dot{\sigma}$ in the sigma domain.

APPENDIX D

Conservation of Physical Properties for Isentropic Surfaces Which Move through the Interface and for Folding Isentropic Surfaces

During the course of a model experiment, isentropic surfaces may rise or sink through the interface resulting

TABLE III. Initialization parameters and diffusion coefficient: θ_s is the potential temperature at earth's surface, K the proportionality constant for the eddy diffusivity.

Experiment	θ_s	$\Delta\theta_s$	γ	$\Delta\gamma$	Longitudinal amplitude	Latitudinal amplitude	Diffusion
E1 & E2	285	5.4	$\theta_s<\theta<315$ K: 5.0×10^{-4}	-0.68×10^{3}	$A=5.0$	$B=2.0$	$K=0.0$ for E1
			$315<\theta<340$ K: 5.0×10^{-4}	$+1.40\times10^{3}$			$K=0.1$ for E2

in a gain or loss of grid points in the isentropic domain. When this occurs, special artificial adjustments must be made to redistribute mass and momentum among grid volumes to insure that mass and momentum are conserved.

If the potential temperature of the solid surface becomes less than the interface potential temperature, a grid point is lost as the surface drops below the interface. In this case, the pressure value stored at that point is set equal to a predetermined value which serves to flag submerged grid points. If the potential temperature difference between the top of a truncated grid volume and the interface exceeds 10 K, a grid point on a solid surface rises through the interface. The initial pressure for the emerging grid point is specified through linear interpolation between the pressures and potential temperatures of the interface and the overlying isentropic surface.

If a grid point on a dashed surface "drops" below the interface (Fig. 16A), the mass and momentum values of the truncated volume predicted at the submerged grid point (1,0) are preserved by adding these quantities to the grid volume (2,0) immediately above by the adjustments

$$(\Delta p)^*_{2,0} = (\Delta p)_{2,0} + (\Delta p)_{1,0}, \tag{27}$$

$$(\Delta pu)^*_{2,0} = (\Delta pu)_{1,0} + (\Delta pu)_{2,0}, \tag{28}$$

which require that

$$u^*_{2,0} = \frac{(\Delta pu)_{1,0} + (\Delta pu)_{2,0}}{(\Delta p)^*_{2,0}}, \tag{29}$$

where the asterisk variable represents an adjusted value, while the non-asterisk variable represents the value predicted immediately before adjustment. An analogous adjustment is made for v^*. After the adjustments are completed, $\Delta p_{(1,0)}$, $u_{(1,0)}$ and $v_{(1,0)}$ are flagged as submerged grid point values. This does not represent a physical vertical transport of mass and momentum through an isentropic surface but a respecification of the physical properties for a grid point $x_{(2,0)}$ now representing an enlarged truncated volume.

When a dashed surface "emerges" (Fig. 16B), physical properties must be redistributed between two grid points in a truncated grid volume in which only the

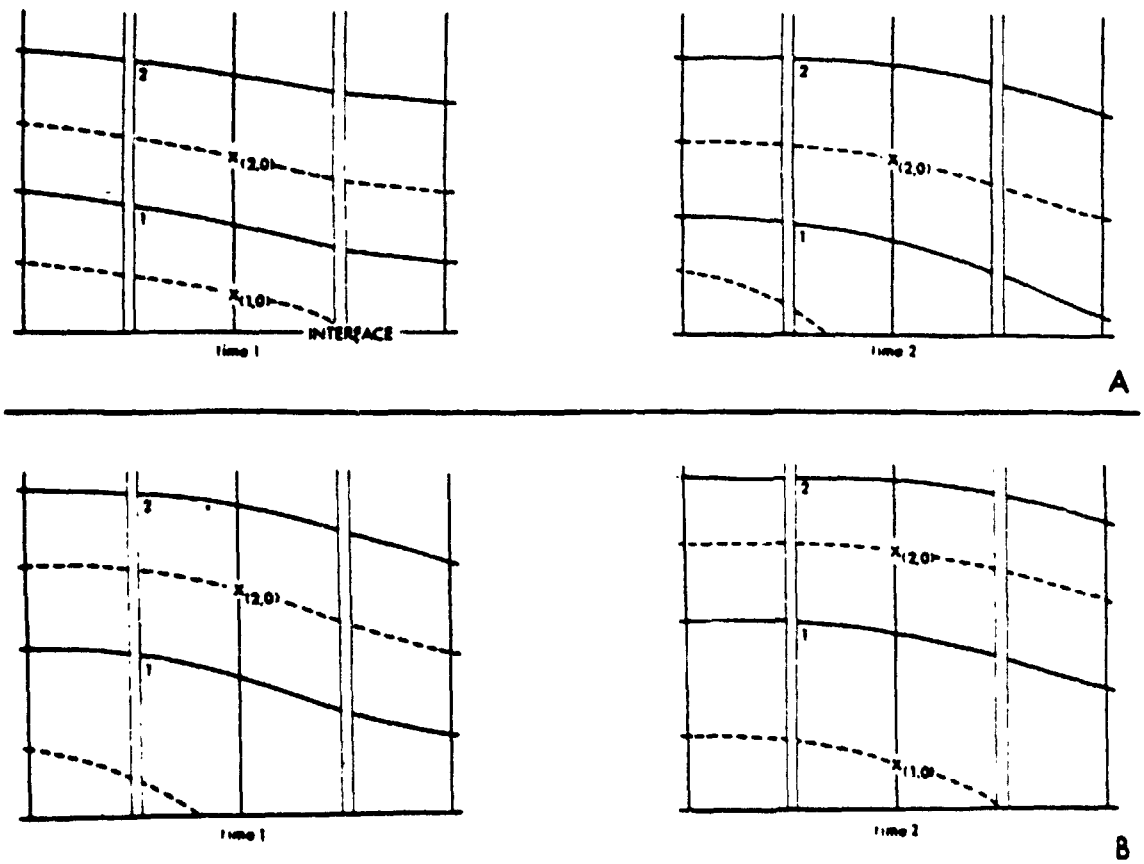


FIG. 16. Grid point $x_{(1,0)}$ on "dashed" isentropic surface representing grid volume 1. (A) submerging into the boundary layer; (B) emerging from the boundary layer. $x_{(2,0)}$ is a grid point representing grid volume 2 immediately above grid volume 1.

grid point (2,0) existed before the emergence of (1,0). The situation is complicated by the need to initialize values of Δp, u and v for three time levels so the predictive equations, approximated by the filtered leapfrog scheme, can be solved at the new grid point. A value for Δp at the emerged grid point (1,0) is determined by a redistribution of mass predicted at the (2,0) grid point (see Fig. 16B). This redistribution, dependent on the ratio of the vertical extent of the new grid volume ($\Delta\theta_1$) to the vertical extent of the entire volume ($\Delta\theta_2$), is accomplished by

$$(\Delta p)^*_{1,0} = (\Delta p)_{2,0}\frac{\Delta\theta_1}{\Delta\theta_2}. \tag{30}$$

The u wind component ($u^*_{1,0}$) is initialized by assigning u_B, the interface value, to the new grid point. To complete the adjustment within the two grid volumes and to insure mass and momentum conservation, new values are assigned to the (2,0) grid point by

$$(\Delta p)^*_{2,0} = (\Delta p)_{2,0} - (\Delta p)^*_{1,0}, \tag{31}$$

$$(\Delta pu)^*_{2,0} = (\Delta pu)_{2,0} - (\Delta pu)^*_{1,0}, \tag{32}$$

which requires that

$$u^*_{2,0} = \frac{(\Delta pu)_{2,0} - (\Delta pu)^*_{1,0}}{(\Delta p)^*_{2,0}}. \tag{33}$$

An analogous approach is used to determine v^*. This procedure is also applied to initialize Δp, u and v at point (1,0) and adjust the parameters at (2,0) for the two previous time increments using the same ratio ($\Delta\theta_1$ $\Delta\theta_2$) throughout. This step simply divides an enlarged truncated grid volume into a regular volume and a smaller truncated volume and does not imply a vertical transport of properties through isentropic surfaces.

Deaven (1976) discussed the need to check for folding isentropes and deep adiabatic layers to preserve the stability of the model integrations. In this model, a limit is set to maintain Δp $\Delta\theta$ within the range 0.1 and 300 mb K^{-1}. If adjustments have to be made to set Δp $\Delta\theta$ within the limits, mass is redistributed in the vertical. In this way the pressure tendency from the interface to earth's surface remains unaffected and generation of spurious oscillations is minimized.

REFERENCES

Asselin, R. A., 1972: Frequency filter for time integrations. *Mon. Wea. Rev.*, **100**, 487–490.

Bjerknes, J., 1951: Extratropical cyclones. *Compendium of Meteorology*, T. F. Malone, Ed., Amer. Meteor. Soc., 577–598.

Bleck, R., 1973: Numerical forecasting experiments based on the conservation of potential vorticity on isentropic surfaces. *J. Appl. Meteor.*, **12**, 737–752.

——, 1974: Short-range prediction in isentropic coordinates with filtered and unfiltered numerical models. *Mon. Wea. Rev.*, **102**, 813–829.

Cahir, J., 1971: Implications of circulations in the vicinity of jet streaks at synoptic scales. Ph.D. thesis, Pennsylvania State University, 170 pp.

Charney, J., 1955: The use of the primitive equations of motion in numerical prediction. *Tellus*, **7**, 22–26.

——, and N. A. Phillips, 1953: Numerical integration of the quasi-geostrophic equations for barotropic and simple baroclinic flows. *J. Meteor.*, **10**, 71–99.

Deaven, D. G., 1976: A solution for boundary problems in isentropic coordinate models. *J. Atmos. Sci.*, **33**, 1702–1713.

Eliassen, A., 1962: On the vertical circulation in frontal zones. *Geofys. Publ.*, **24**, 147–160.

——, and F. Raustein, 1968: A numerical integration experiment with a model atmosphere based on isentropic coordinates. *Meteor. Ann.*, **5**, 45–63.

——, and ——, 1970: A numerical integration experiment with a six-level atmospheric model with isentropic information surface. *Meteor. Ann.*, **5**, 429–449.

Friend, A. L., 1976: Numerical weather predictions using a combination of isentropic and sigma coordinates. Ph.D. thesis, Texas A&M University, 69 pp.

——, D. Djuric and K. C. Brundidge, 1977: A combination of isentropic and sigma coordinates in numerical weather prediction. *Beitr. Phys. Atmos.*, **50**, 290–295.

Gall, R. L., 1972: Prediction of a quasi-steady propagating jet core with an isentropic numerical model. Ph.D. thesis, University of Wisconsin, 121 pp.

Haltiner, G. J., and J. M. McCollough, 1975: Experiments in the initialization of a global primitive equation model. *J. Appl. Meteor.*, **14**, 281–288.

Hoskins, B. J., and F. P. Bretherton, 1972: Atmospheric frontogenesis models. Mathematical formulation and solution. *J. Atmos. Sci.*, **29**, 11–37.

Johnson, D. R., 1970: The available potential energy of storms. *J. Atmos. Sci.*, **27**, 727–741.

——, 1977: Generalized transport relations for parametric grid volumes. Isentropic numerical models: Results on model development for zonally averaged and secondary circulations. Sci. Rep. to NSF, Dept. of Meteorology and Space Science and Engineering Center, University of Wisconsin-Madison, 279–273 [NTIS No. PB 283480/AS].

Kurihara, Y. and J. L. Holloway, 1967: Numerical integration of a nine-level global primitive equation model formulated by the box method. *Mon. Wea. Rev.*, **95**, 509–529.

Lorenz, E. N., 1955: Available potential energy and the maintenance of the general circulation. *Tellus*, **7**, 157–167.

Murray, R., and S. M. Daniels, 1953: Transverse flow at entrance and exit to jet streams. *Quart. J. Roy. Meteor. Soc.*, **79**, 236–241.

Namias, J., and P. F. Clapp, 1949: Confluence theory of the high tropospheric jet stream. *J. Meteor.*, **6**, 330–336.

Newton, C. W., 1959: Axial velocity streaks in the jet stream ageostrophic "inertial" oscillations. *J. Meteor.*, **16**, 638–645.

Phillips, N. A., 1957: A coordinate system having some special advantage for numerical forecasting. *J. Meteor.*, **14**, 184–185.

——, 1960: On the problem of initial data for the primitive equations. *Tellus*, **12**, 121–126.

Reiter, E. R., 1963: *Jet Stream Meteorology*. The University of Chicago Press, Chaps. 4, 6.

——, 1967: *Jet Streams. How Do They Affect Our Weather*. Anchor Books, Doubleday, 189 pp.

——, 1969: Tropopause circulation and jet streams. *World Survey of Climatology: Climate of the Free Atmosphere*, Vol. 4, D. F. Rex, Ed., 85–193.

Richtmyer, R. D., and K. W. Morton, 1967: *Difference Methods for Initial Value Problems*, 2nd ed. Interscience-Wiley, Chaps. 7, 8.

Riehl, H., *et al.*, 1952: Forecasting in the middle latitudes. *Meteor. Monographs*, No. 5, Amer. Meteor. Soc., 80 pp.

Sawyer, J. S., 1956: The vertical circulation at meteorological

fronts and its relation to frontogenesis. *Proc. Roy. Soc. London*, A234, 346–362.

Shapiro, M. A., 1974: The use of isentropic coordinates in the formulation of objective analysis and numerical prediction models. *Atmosphere*, 12, 10–17.

——, 1975: Simulation of upper-level frontogenesis with a 20-level isentropic coordinate primitive equation model. *Mon. Wea. Rev.*, 103, 591–604.

Smagorinsky, J., 1958: On the numerical integration of the primitive equations of motion for baroclinic flow in a closed region. *Mon. Wea. Rev.*, 86, 457–466.

Starr, V. P., 1945: A quasi-Lagrangian system of hydrodynamical equations. *J. Meteor.*, 2, 227–237.

Trevisan, A., 1976: Numerical experiments on the influence of orography on cyclone formation with an isentropic primitive equation model. *J. Atmos. Sci.*, 33, 768–780.

Uccellini, L. W., D. R. Johnson and R. E. Schlesinger, 1977: An isentropic and sigma coordinate hybrid model. *Isentropic numerical models: Results on model development for zonally averaged and secondary circulations*, Sci. Rep. to NSF, Dept. of Meteorology and Space Science and Engineering Center, University of Wisconsin-Madison, 179–254 [NTIS PB 283480/AS].

University of Chicago, Dept. of Meteorology, 1947: On the general circulation of the atmosphere in middle latitudes. *Bull. Amer. Meteor. Soc.*, 28, 255–280.

Whittaker, T. M., 1977: Automated streamline analysis. *Mon. Wea. Rev.*, 105, 786–788.

——, and R. A. Petersen, 1977: Objective cross section analysis incorporating thermal enhancement of observed winds. *Mon. Wea. Rev.*, 105, 147–153.

Paper 58

Reprinted from MONTHLY WEATHER REVIEW, Vol. 107, No. 12, December 1979
American Meteorological Society
Printed in U. S. A.

Shallow Convection on Day 261 of GATE: Mesoscale Arcs

C. WARNER AND J. SIMPSON[1]

University of Virginia, Charlottesville 22903

D. W. MARTIN, D. SUCHMAN AND F. R. MOSHER

University of Wisconsin, Madison 53706

R. F. REINKING

Boundary Layer Dynamics Group, NOAA/ERL, Boulder, CO 80303

(Manuscript received 11 June 1979, in final form 5 October 1979)

ABSTRACT

On 18 September 1974, a cloud cluster growing in the GATE ship array was examined using aircraft flying close to one another at different heights, the geostationary satellite SMS-1, and radar, rawinsonde and ship data, with a view to elucidating mechanisms of convection. In this paper we concentrate analysis on cloudy convection in the moist layer.

In and above southerly surface monsoon flow approaching the cluster, clouds indigenous to the moist layer took the form of rows of tiny cumulus, and of arcs of cumulus mediocris, with patterns different from those of deeper clouds. From satellite visible images, arcs were traced for periods exceeding 2 h. Airborne photography showed that the arcs were composed of many small clouds. Radar data showed that they originated after precipitation. Apparently, throughout their life cycle, they perpetuated the pattern of an initiating dense downdraft. Eventually they yielded isolated cumulus congestus, again bearing precipitation. Aircraft recorded the distribution of thermodynamic quantities and winds at altitudes within the mixed layer, and at 537 and 1067 m. These data indicated that the arcs persisted as mesoscale circulations driven by release of latent heat in the clouds, rather than being driven by the original density current at the surface. The cloudy circulations were vigorous near and above cloud base, becoming weaker upward through altitude 1 km. The entire mesoscale circulation systems were of horizontal scale roughly 40 km.

The mesoscale cloud patterns of the moist layer appeared to play a primary role in heat transfer upward within this layer, and contributed to the forcing of showering midtropospheric clouds.

1. Introduction

The Atlantic Tropical Experiment of the Global Atmospheric Research Program (GATE) was centered on an area near 8.5°N, 22.5°W during summer 1974. On 18 September 1974 (day 261) a growing cloud cluster was monitored using the satellite SMS-1, ships and aircraft. A preliminary synthesis of observations of this cluster was given by Warner *et al.* (1977). Suchman and Martin (1976) included data from day 261 in an assessment of wind measurements by satellite tracking of clouds, and Suchman *et al.* (1977) have given a preliminary description of the evolving circulation of the cluster. Warner and Austin (1978) have described statistical aspects of radar echoes of the cluster. As will be seen in Section 4, much of the shallow convection associated with the cluster was organized in a manner very different from the deep convection. It is the purpose here to describe the shallow convection in the lowest 2–3 km of the moist layer south of the cluster, which contained the surface monsoon air flowing northward toward the cluster. Warner *et al.* (1980) treat the deep convection. These papers are condensed from a detailed report (Warner *et al.*, 1978). A future treatment will show how the cluster, occurring in relatively undisturbed conditions, was related to activity on larger scales.

A major GATE objective is to assess the interactions of air motions on different scales. This paper documents interactions between clouds indigenous to the moist layer and clouds of the mid-troposphere, and describes related events in the subcloud layer. An attempt is made to assess the role of the moist layer clouds in vertical exchanges of energy. The paper is chiefly concerned with arcs, the reader may like first to inspect the final figure—a schematic.

[1] Present affiliation: Goddard Laboratory for Atmospheric Science, NASA, Greenbelt, MD 20771.

0027-0644/79/121617-19$08.75

TABLE 1. Times (GMT) of passage of the aircraft at the corners of the box circuit, defined by latitudes 8°34′ and 9°56′N, and longitudes 21°2′ and 22°23′W. (The north leg was flown from west to east, then east to west, in succession.)

Aircraft	Altitude (m)	Corner			
		D_1 (NE)	D_2 (SE)	D_3 (SW)	D_4 (NW)
Sabreliner (NCAR)	12500	1354	1607	1618	1631
US-C130 (NOAA)	5800	1341 1505–12 1721	1400–03 1701–1645	1422–24 1555–52	1444–47 1553–31
DC-6 (NOAA)	1067	1343 1528–32 1727	1409–11 1702–1659	1436 1635	1502–04 1608–1558
Electra (NCAR)	537 & 35	1342 1524–28 1717	1406–10 1654–52	1433–37 1623–19	1459–1502 1555–52
UK-C130	238 & 52	1358 1549–53 1751	1421–25 1729–25	1447–1501 1703–1656	1524–27 1633–16

Black (1978) shows satellite photographs of two arc patterns which he associates with large cumulus clouds. That in Black's Figs. 9 and 10 may have been similar in nature to the arcs described here. Black calls for coordination of satellite and ground-based data. Here we provide this, and obtain a view of the arcs as self-perpetuating mesoscale entities, arising from and bearing the pattern of an initiating cold air outflow concomitant with rainfall, and eventually yielding isolated cumulus congestus, again bearing rain.

After a brief description of sources of data in Section 2, a representative aerological sounding is shown in Section 3, and maps of both shallow and deep clouds in Section 4. Arc patterns in the moist layer are traced in Section 5; they showed persistence over 2 h. Aircraft traverses through arcs are examined in Section 6, aircraft soundings and surface conditions in Section 7 and turbulent fluxes in Section 8, to throw light on how the arcs persisted and on the nature of the air circulations associated with them. A summary is given in Section 9 and conclusions are drawn in Section 10.

2. Sources of data

a. Aircraft data

Five aircraft simultaneously boxed a square of side 150 km which encompassed towers of the cluster. The different aircraft and flight altitudes, and the times of passing the box corners, are given in Table 1. The first box circuit was accomplished between 1340 and 1550 GMT by three aircraft in the low-level moist layer (up to 2.3 km) and one at ~6 km, the second between 1550 and 1750 GMT with an additional aircraft at altitude 12.5 km. The aircraft measured thermodynamic and wind data. All but the lowest carried time-lapse movie cameras; cloud maps were made using the methods of Ronne (1959) and Warner (1978). Turbulent heat fluxes were obtained from a gust probe on the DC-6 (Bean *et al.*, 1976).

b. Satellite data

From the geostationary satellite SMS-1, visible and infrared images with resolution ~1 km and 8 km, respectively, were obtained every 15 min, and used for following cloud developments and obtaining winds by cloud tracking (Suchman and Martin, 1976).

c. Radar data

Precipitation echoes were recorded digitally by the 5 cm radar aboard the Canadian ship *Quadra*, situated near 9°3′N, 22°36.5′W. This radar, and the data, have been described by Warner and Austin (1978). A three-dimensional scan cycle was completed every 5 min except during the period 1448–1614 GMT.

d. Rawinsonde data

Soundings from ships of the GATE B and A/B scale arrays have been used in this study, in their original form as supplied by the Center for Experiment Design and Data Analysis (NOAA).

e. Ship data

Thermodynamic and wind data from the booms of the ships *Dallas* (at 8°32′N, 22°31′W) and *Oceanographer* (at 7°44′N, 22°14′W) were compared with the all-sky movies taken aboard these ships. Profiles from the *Dallas* boundary layer instrument system [(BLIS) Garstang *et al.*, 1977] were examined.

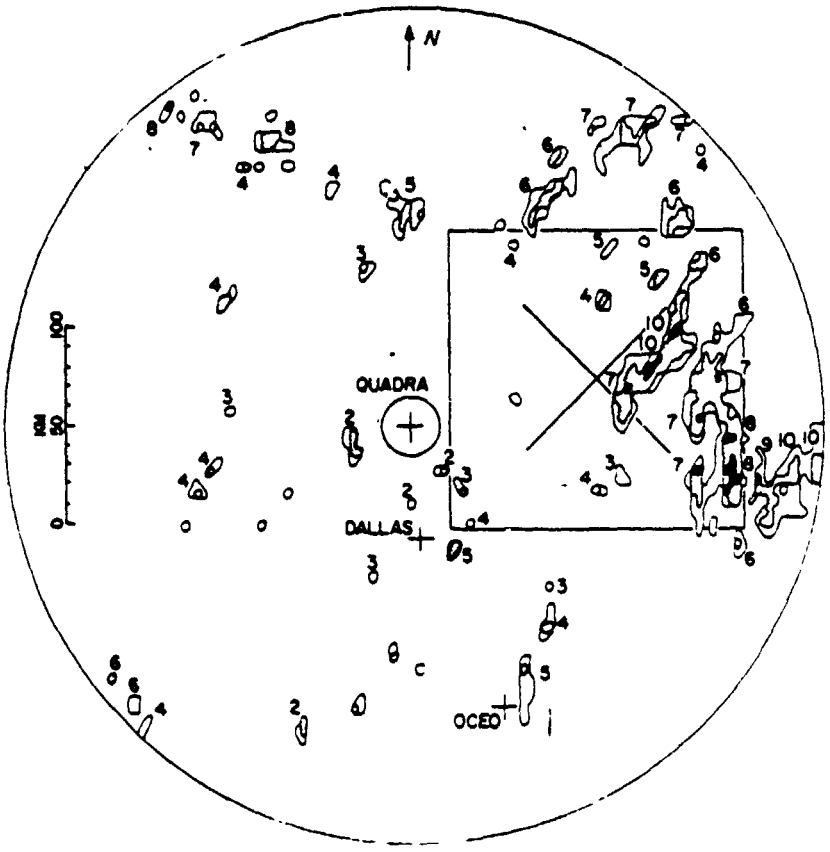


FIG. 1. 1629 GMT. Echoes from the *Quadra* radar, situated near 9°3'N, 22°36 5'W. Antenna elevation 0.35°. Maximum range 210 km. Minimum range 20 km (limited by clutter from the sea). Contours of reflectivity at the thresholds 16, 24 and 35 dB above 1 $mm^6 m^{-3}$ (dBZ)—equivalent to rainfall at about 0.3, 1.3 and 9 mm h^{-1}, measured by averaging over bins of area 18 km^2. The box was that flown by the aircraft. The cloud cluster was growing in the east half of the box. Digits refer to heights (km) of echo cores of reflectivity reaching 29 dBZ.

a radar view of the cluster, showing the positions of the aircraft circuit and of the ships, is given in Fig. 1.

3. A rawinsonde

The 1500 GMT *Oceanographer* sounding, shown in Fig. 2, appears to have been representative of undisturbed conditions southwest of the growing cluster, which was propagating slowly southwestward at ~2.5 m s^{-1}. In the sounding, several layers characteristic of the marine tropical atmosphere (Riehl, 1954) can be identified. The moist layer reached ~2.2 km. It contained a mixed layer of depth roughly 500 m. Surface winds at the *Oceanographer* reached 6–7 m s^{-1} from the south. The moist layer was capped by a stable layer, where the air was very dry and winds were light and from the northeast or east.

At ~4.5 km (600 hPa), the air was moist. In this humid layer the winds were northerly at ~6 m s^{-1}, and the stratification was conditionally unstable. There was a stable layer at 5 km. (There were many stratus and stratocumulus patches at ~2.5 and 5 km.)

Above 6 km the air was dry, with easterlies that persisted up to 8 km, at 7 m s^{-1} in this dry layer.

4. Cloud maps from the first circuit

The cloud maps shown in Fig. 3 were obtained using a 35 mm side camera on the US-C130 (with black and white film), and a 16 mm camera in the nose. The aircraft flew at ~130 m s^{-1}, and photographs were taken every 5 s (or ~650 m). Mapping was based chiefly on the C130 right side camera, and followed the method described by Ronne (1959). For an aircraft flying steadily past a cloud with a side-pointing cine-camera of constant frame rate, the range of the cloud is nearly proportional to the time or the number of frames of film required for it to pass across the field of view. If the position of

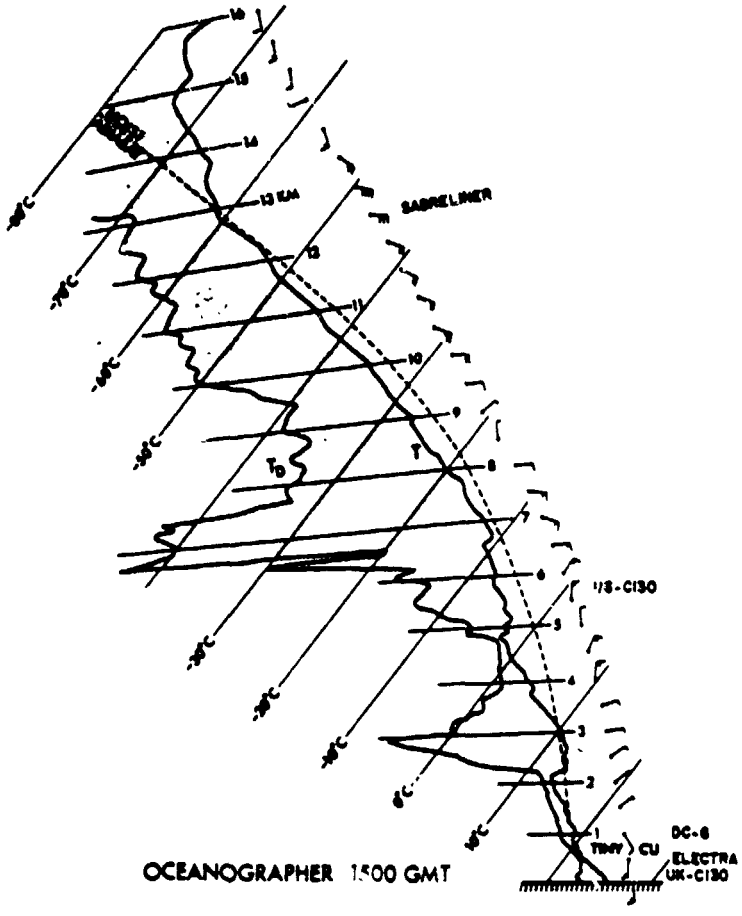


FIG. 2. 1500 GMT. Tephigram from *Oceanographer* situated at 7°44'N, 22°12'W. Solid lines: dew point (T_D) and temperature (T); dashed line: moist adiabat (of constant Rossby's θ_E). Vertical scale is height (km) above the sea. Flight levels of the aircraft are indicated. Winds are shown every 500 m, with one barb = 5 m s^{-1}. Level of tiny cumulus cloud is indicated.

the aircraft is known for any arbitrarily chosen frame, then the cloud position may be plotted on a plan view map. The position of the aircraft was known to a nominal precision of 2–4 km by inertial navigation. The plotting technique was of comparable accuracy (Snow *et al.*, 1975). To find the height of a cloud element, its elevation was measured relative to the horizon; its height then followed from the range. Cloud heights obtained from the C130 side camera agreed to within a few hundred meters with heights derived from an algorithm of Mosher (1975), which yields the height of a cumulus top as a function of visible and infrared brightness seen by satellite. Comparisons are detailed in Table 2. Nominal accuracies in the maps are a few kilometers in the horizontal and ~500 m in the vertical.

At 1437 GMT, while the aircraft were near the southwest corner of the box, the satellite recorded the pattern of visible brightness shown in Fig. 4. Across the east side of the box a cirrus shield was prominent. Isolated bright patches were mostly groups of cumulus, in the southerly anticyclonic flow in the moist layer (Suchman *et al.*, 1977).

The first circuit by the US-C130 was made between about 1341 and 1445 GMT. Prior to this, the aircraft flew across the box, crossing the middle at 1305. The resulting cloud map is presented in two parts. Fig. 3a shows active cumulus tops with heights in the interval 2.5 km to the maximum, exceeding 13 km in the northeast corner of the box. Fig. 3b shows tops below 2.5 km, and the horizontal limits of cirrus anvils. The picture is not complete, particularly of the small clouds, because they were very numerous, and photographic coverage was not obtained simultaneously over the whole area.

Fig. 3a shows relatively few clouds in the southwest part of the box, and more to the northeast, with greater heights. At the time, the cloud cluster was growing in the northeast corner, and propagat-

ing slowly into the box. A cloud with top exceeding 13 km was observed by satellite near D_1. Most cumulus clouds taller than ~3 km tended to be aligned in bands of orientation about 215°, roughly 9 km apart.

Fig. 3b shows that low clouds were arranged differently from the cumulus and cumulonimbus clouds shown in Fig. 3a. A region to the northeast was occupied by bands of altocumulus at 2.9 km, of orientation about 190° and spacing ~3 km.

In the west part of the box, many of the small clouds were arranged in arcs. They were composed of cumulus mediocris sometimes reaching roughly the top of the moist layer (2–3 km). One such arc, obscured from satellite view by a cirrus layer over the north side of the box, has been documented by Warner (1978); this arc is shown in Fig. 5. The clouds leaned backward, in a manner corresponding with winds decreasing with height upward from the top of the mixed layer (Fig. 2). Low-level winds were toward the inside of the arcs.

At the lowest end of the scale of convective motions, there were many tiny cumulus fractus, of linear dimension ~500 m, with bases at about 500–700 m. These were aligned in rows oriented north–south, with separations of 1 or 2 km along and between rows. The rows were aligned approximately with the surface wind. Tiny cumulus also appeared in front of the arcs (Fig. 5).

In connection with the tiny cumulus, Fitzjarrald (1978) has used the *Oceanographer* acoustic sounder data to examine the dominant horizontal scale (l) of free convection in the moist layer. For the period 0827–1225 GMT on day 261, he gave the dominant local period of fluctuation as 6.25 min, the wind speed as 7 m s^{-1}, the depth of the mixed layer (Z_i) as 450 m, and the Monin-Obukhov length (L) as −119 m. The first two numbers yield l = 2.6 km, the second that $-Z_i/L$ = 3.8. The nondimensional horizontal wavelength l/Z_i = 5.8. His Fig. 7 then indicates that this constitutes a relatively low instability, with a relatively large horizontal wavelength. The number 2.6 km corresponds quite well with the spacing of the tiny cumulus along rows shown by the mapping. The information on the rows of tiny cumulus is consistent with the finding of LeMone and Pennell (1976, Fig. 16a) that the rows resulted from forcing below cloud base associated with longitudinal roll circulations, oriented roughly along the wind direction. Tiny cumulus in front of the arc shown in Fig. 5 were aligned parallel to the arc, apparently transverse to the wind direction rather than along it.

With many active towers reaching stable layers, much stratus was present (a feature seen less often in the undisturbed Caribbean). The deep clouds produced anvils at about 13 km (Fig. 3b). Their leading edges spread from slightly south of east at

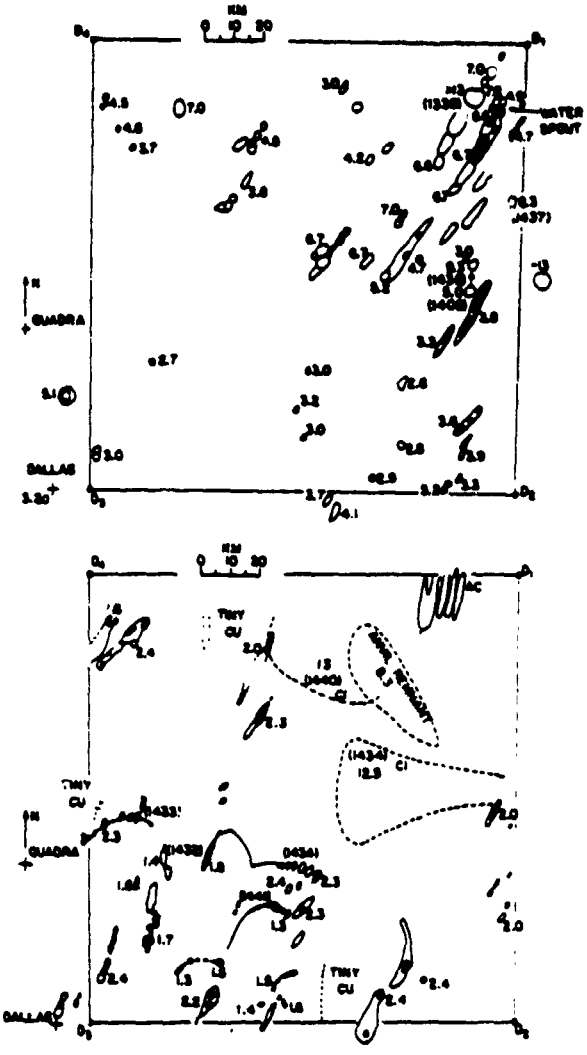


FIG. 3. 1300 to 1445 GMT. a (top): Active tops (outlined by contours) in the height interval 2.5 to >13 km. b (bottom): Clouds of height <2.5 km, and anvils (dashed). Thin solid lines represent arcs of clouds. Numbers are heights (km) above the sea, and times of measurement. The corners D_1 to D_4 define the box circuit (see Table 1).

~25 m s^{-1}. Towers reaching different levels in the troposphere produced individual anvil plumes. An isolated remnant of anvil at 8.5 km was oriented along 145° (rather than ~90° as at 13 km). Many stratus and stratocumulus patches at ~2.5 and 5 km are not portrayed here.

5. Mesoscale patterns seen by satellite, radar and aircraft photography

A 16 mm movie loop of visible satellite images of the boxed area has been made, with time intervals of 15 min. In this movie, it is possible to follow several arc-shaped mesoscale patterns over periods

TABLE 2. Cloud tops by photogrammetry, and cloud tops (and bases) from satellite data by Mosher's (1975) method. (McIDAS is the man-computer interactive data access system at the University of Wisconsin.)

Cloud map position		Height (km)	Time (GMT)	McIDAS position		Top (base)	Time	Comment
8°36'N	21°22'W	~2.5	1407			2.1 (0.2)		Inactive Ac
8°54'	21°23.5'	2.6	1408			1.8 (0.3)	1407	Isolated Cu at end of dissipated band
				8°51.3'N	21°24.9'W	1.9	1422	
9°6.5'	21°10' to 21°11'	3.8	1352			3.2 (0.6)		High top on Cu band
				9°20.5'	21°34'	4.6 (2.9)	1322	
						4.4 (3.2)	1337	Motion: 3.7 m s^{-1}/030°
9°13'	21°32' to 21°33'	6.1	1351					Anvil remnant
				9°10'	21°40'		1422	Dissipated
				9°33'	21°9'	10.5	1137	
				9°33'	21°13'	10.0	1152	
				9°34'	21°16'	9.9	1207	
				9°35'	21°17'	9.4	1222	Motion: 4 m s^{-1}/030°
				9°33'	21°15'	8.9	1237	
9°37'	21°28'	8.3	1348					Anvil remnant
9°14.5'	21°3'	13	1349	9°12.1'	21°3.2'	12.2	1352	
9°47'	21°18'	13.1	1336			9.7 (2.5)	1337	
						12.7	1352	
8°58'	21°4'	15	1621	9°16'	21°23'	15.4	1807	Highest towers by each method

exceeding an hour. Three such patterns are shown Figs. 6a–6c. The square box was slightly distorted in the movie, to the shape shown in the figure. It is possible to pick out the positions of these patterns at 1437 in Fig. 4, the visible satellite photograph for that time. The arcs were composed essentially of small clouds, as represented diagrammatically in Fig. 3b.

A frame from the US-C130 nose camera movie at 1428:35 GMT is shown in Fig. 7. The aircraft

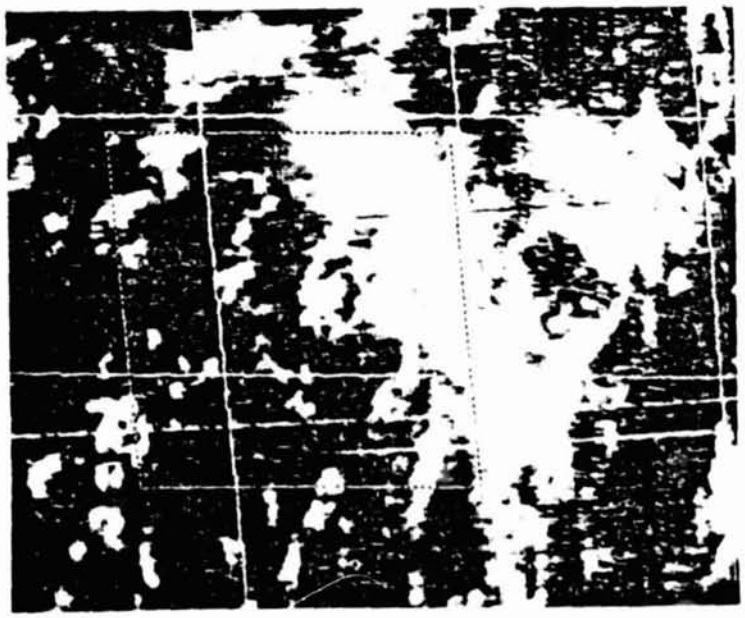


FIG. 4. 1437 GMT. Visible satellite image. The cloud cluster is seen at right center, with anvil cirrus streaming northwestward, north being up the page. The box circuit is shown by pecked lines; a side was 150 km long. Light fiducial lines show the latitudes 9 and 10°N, and longitudes 20, 21 and 22°W.

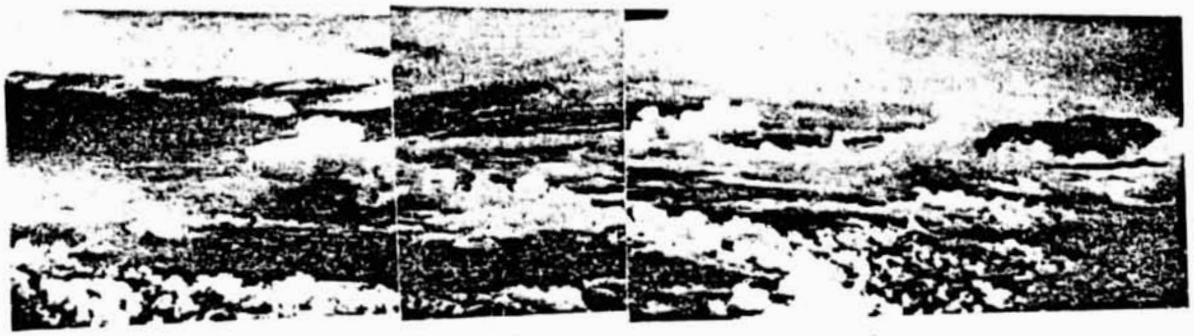


FIG. 5. 1522 GMT. Composite of photographs looking north from the US-C130 aircraft flying at altitude 5.7 km on the north side of the box circuit. High towers of the cloud cluster were to the east and south of the aircraft. Anvil cirrus from them was overhead. In the right foreground is an arc of clouds, moving from roughly the west-southwest (toward the right). Cumulus mediocris in the arc were of maximum height about 2.7 km. They leaned backward. A high tower (at left) reached 9 km. It apparently originated from the arc. Tiny cumulus preceded the arc. Note the evidence of organization in their pattern. Major features have been mapped in Warner (1978).

was flying northward at altitude 6.5 km, coming up on the arc shown in Fig. 6c. The faint diffuse line seen in the satellite view (Fig. 4) is revealed as an arc of cumulus mediocris, preceded by tiny cumulus. From the satellite movie, it was evident that while the arc patterns propagated consistently, their composition continuously changed, with small clouds dissipating and being replaced by new clouds, of lifetime much less than the pattern. Near the west side of the box, the velocity of propagation was from about 198° at 6 m s^{-1}, matching the wind near cloud base. Cloud bases were at 500 m (near the top of the mixed layer) and higher. The pattern velocities measured from the satellite movie loop show that previous estimates from aircraft photogrammetry (Warner *et al.*, 1977) were much too great.

The origins of the arcs and their relationships to

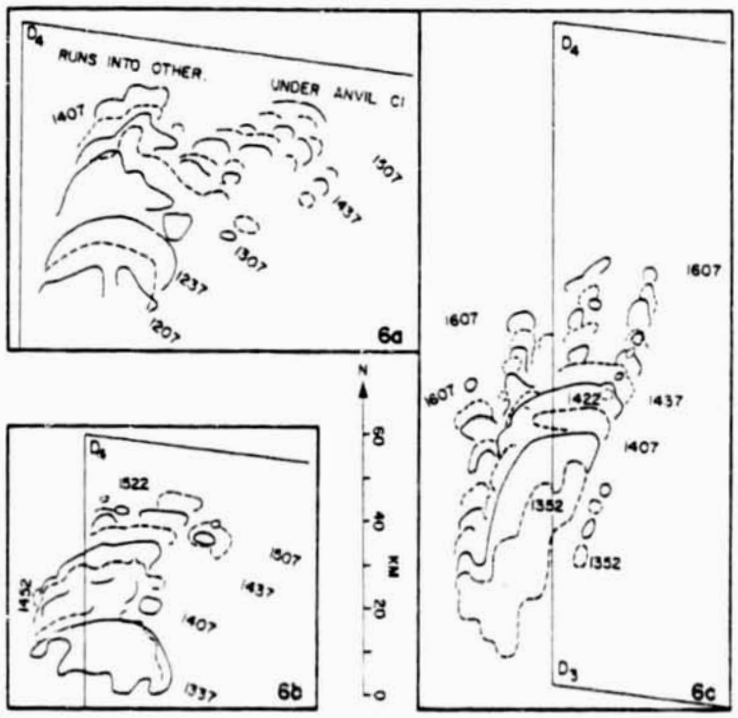


FIG. 6. Motions of arcs traced from satellite movie. The west side D_3 D_4 of the aircraft box circuit is shown. Defined by the latitudes 8°34' and 9°56'N, and longitudes 21°2' and 22°23'W, the box was slightly distorted from square shape in the movie.

FIG. 7. 1428:35 GMT. Frame from nose camera movie of US-C130 aircraft, taken at altitude 6.5 km while flying northward on the first circuit, coming up on the arc shown in Figs. 6c and 9.

other features were examined. Contour analyses were drawn by hand at an easily discerned threshold of brightness, close to the edge of the visible cloud, seen in projected frames of the movie loops. These maps, at quarter-hour intervals from 1205 to 1435 GMT, appear in Figs. 8 and 9. Superimposed on the contour analyses are alphanumerics representing radar reflectivity, transcribed manually from radar maps as illustrated in Fig. 1 of Warner and Austin (1978). These represent averages over 18 km^2. The relationships between alphanumerics, reflectivity Z and rainfall rate R are shown in Table 3 (for three different expressions for Z, as indicated, the middle column in the table is preferred).

In Figs. 8 and 9, attention is drawn to the arcs in Figs. 6a and 6c, respectively, by small arrows indicating the directions of pattern motion. Occasionally dashed lines appear, where the leading edge of the arcs was marked by faint rather than bright clouds as seen from the satellite. From echo top plan position indicator maps (Warner and Austin, 1978), occasional digits within small squares show peak heights (km) of echo cores. Here a "core" exceeded reflectivity 29 dBZ and occurred within groups of echoes associated with the arcs.

The radar echoes are quite well correlated with the brightness contours. The correlation is not perfect; this is to be expected, because once falling raindrops of size exceeding 0.5 mm and with number densities large enough to yield a radar echo have appeared, small cloud droplet populations are depleted by coalescence or evaporation. It seems from Figs. 8 and 9 that the area of radar averaging (18 km^2) is rather large. Both satellite brightness maps and the radar alphanumerics give rather a coarse view of the activity.

Considering first the arc shown in Figs. 6a and 8, a close association with echoes was apparent until 1235. At 1305 there was bright cloud, but few echoes. More echoes appeared after 1335, with cores often reaching an altitude of 6 km. The persistent arcs were a favored location for generation of new echoes.

The system shown in Figs. 6b and 8 appeared just as a short-lived group of echoes was dissipating. The mesoscale patterns appear to have been generated in association with echoes, and subsequently to have given rise to new clouds and echoes. The same applies to the pattern shown in Figs. 6c and 9.

Attention is drawn to a ring of mean diameter roughly 70 km discernible in Fig. 8 between 1305 and 1405, and straddling the north side of the box. (The north side roughly bisected this feature at 1335.) The arcs of Fig. 6a were on its southwest border. It moved north-northeastward with the low-level flow.

Finally, it may be seen that echoes persisted intermittently near the *Dallas* during the period of Fig. 9. Echo cores reached 4 km. Cloud tops probably reached 5 or 6 km where winds were from the north, rather than from the south as in the mixed layer. Winds averaged through the thickness of such deep clouds were near zero; this probably accounts for the persistence.

The above observations imply that the mesoscale arcs were self-regulating, in a layer near the surface in which winds were from the south-southwest. This corresponds roughly to the lowest kilometer. Precipitation cores extending into the lower middle troposphere apparently generated new mesoscale patterns, and new cores in turn were generated by them.

We may now look from the satellite at a relatively large area, roughly an order of magnitude greater than the area of Fig. 4. Figs. 10a–10c are photographs of intermediate scale at 1422, 1637 and 1825 GMT, respectively. A separate drawing shows corresponding points of reference. Relatively clear air north and west of the cloud cluster is readily seen in Figs. 10; the cluster is toward the lower right in each picture, with cirrus streamers extending northwestward. West and southwest of the cluster, the bands of lower tropospheric clouds represented in Figs. 3a and 4 may be seen. A closer look is necessary to see the lesser clouds like those shown in Figs. 3b, 6 and 7. Much fine tracery is apparent. At 1637 and 1822 arcs were numerous, convex toward roughly the north or northeast, in the direction of winds near the surface. Their radius of curvature was about 20 km, and scale size roughly 40 km.

A comparison of the series of prints in Fig. 10 with the cloud motion fields at 1330 and 1500 GMT by Suchman and Martin (1976), and at 1130 and 1500 by Chatters and Norton (1977), suggested tentatively that the arcs appeared within the moist southerly flow when this exceeded a certain speed, roughly 5 m s^{-1}.

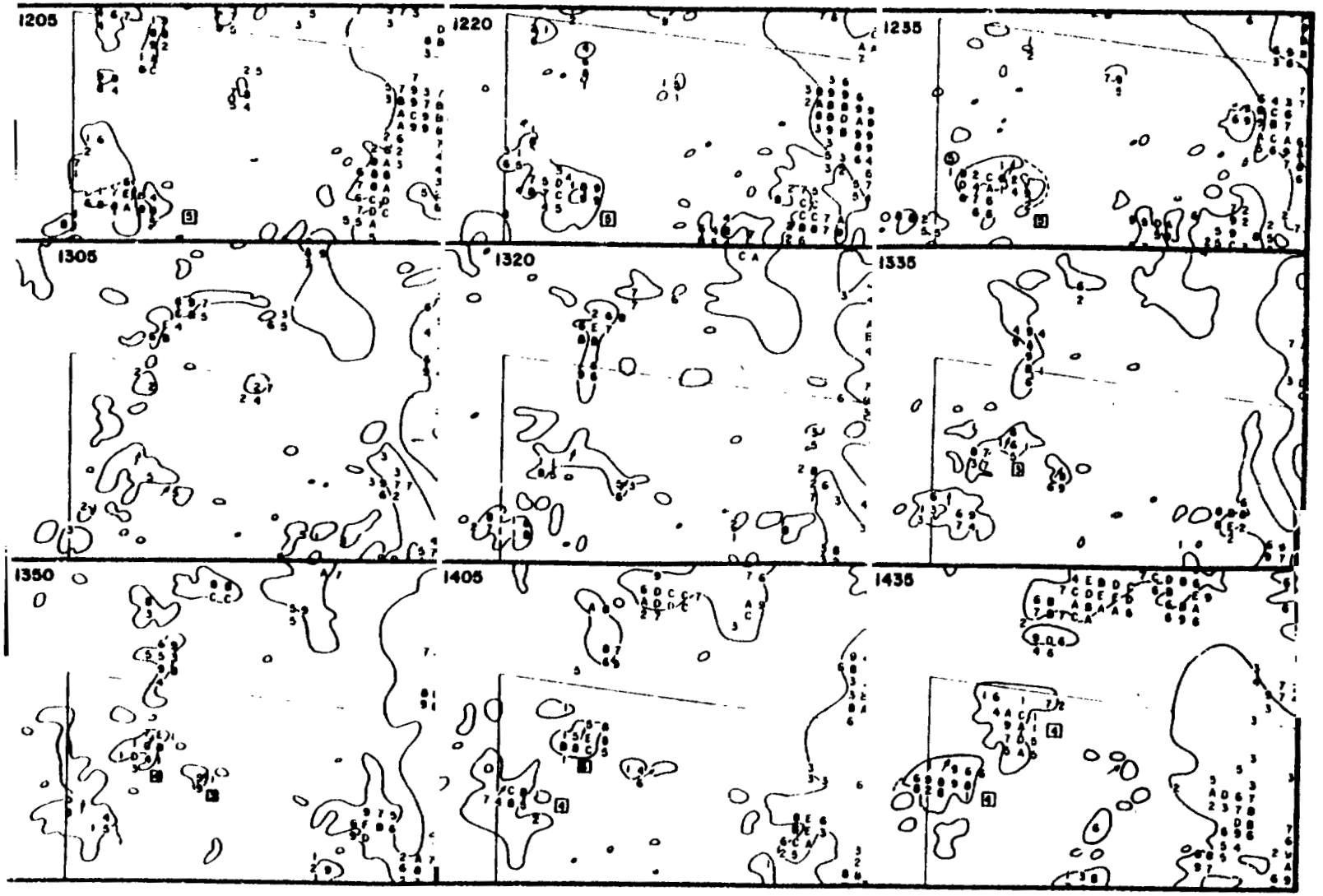


FIG. 8 Maps of visible satellite brightness, together with radar reflectivities in bins measuring 4.2 km north-south by 4.2 km east-west, represented by alphanumerics (see Table 3). Successive times are from 1205 to 1435 GMT at 15 min intervals (two being missing). The satellite and radar scan times differed by 4 min. Radar antenna elevation 0.3°. From echo top plan position indicator maps (Warner and Austin, 1978), occasional digits within small squares show peak heights (km) of echo cores, exceeding reflectivity 29 dBZ, within groups of echoes associated with the arcs. A part of the aircraft box circuit appears in each panel, as in Fig. 6. Solid arrows and dashed lines (indicating a boundary of low brightness) refer to the arc patterns shown in Figs. 6a and 6b.

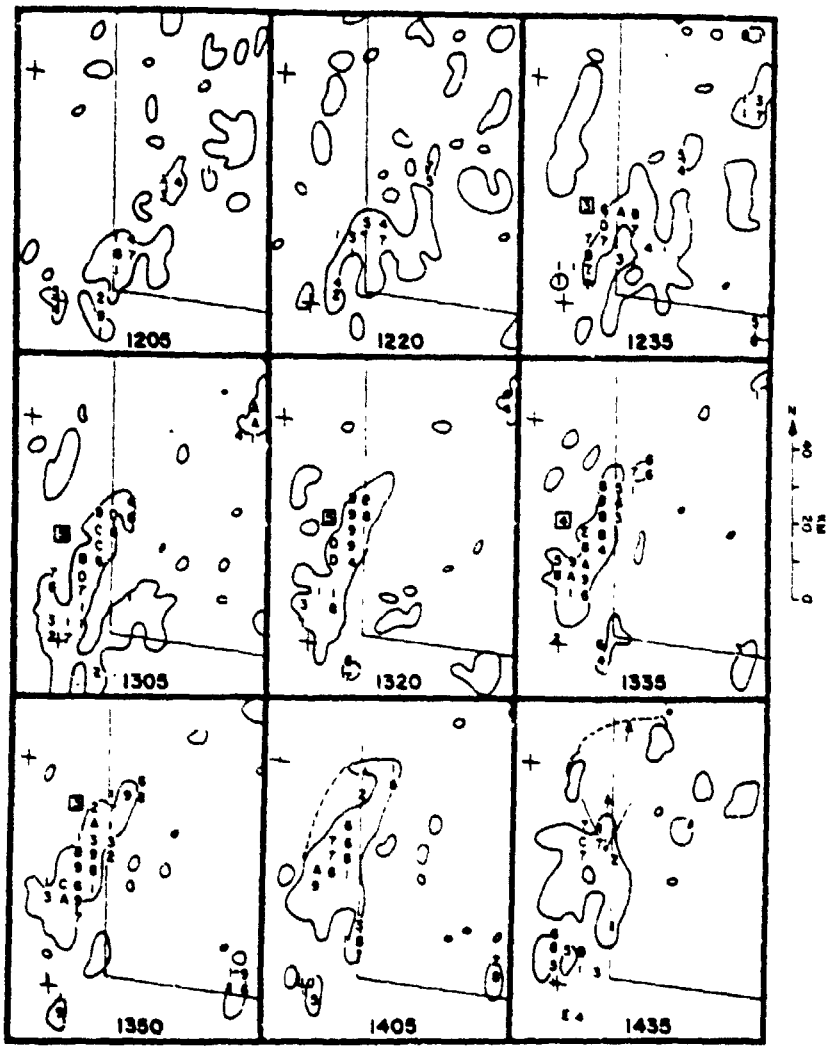


FIG. 9. Maps of visible satellite brightness and radar reflectivities as in Fig. 8. Crosses show the locations of the ships *Quadra* (at 9°3 N, 22°36 5'W) and *Dallas* (at 8°32'N, 22°31'W). Solid arrows and dashed lines refer to the arc pattern shown in Fig. 6c. The photograph Fig. 7 was taken at 1428:35 from an altitude of 6 5 km, the position and field of view are shown with an open arrow in the panel for 1435.

6. Aircraft traverses through mesoscale arc patterns

In this section aircraft data are displayed, which can partly explain how the arcs persisted for up to 2 h, while being composed of elements of individually much shorter lifetimes.

The UK-C130 aircraft flew within the mixed layer, alternately at nominal altitudes of 238 and 52 m, or pressures of 985 and 1007 hPa. The Electra flew just above the mixed layer at nominal altitude 537 m (949 hPa), the DC-6 at 1067 m (891 hPa). The Electra had side-pointing movie cameras with a frame interval of 4 s (~430 m); the DC-6 had a nose camera with a frame interval of 5 s (490 m). From the electra movies, frames were noted for which the aircraft was flying through either clear air, fields of tiny cumulus fractus of linear dimension ~500 m, or fields of cumulus mediocris, or was penetrating a cloud. Using the nose camera on the DC-6, it was possible to do the same analysis, and also to discern the presence of rain.

The DC-6, Electra and UK-C130 flew close to one another in the moist layer during the box circuits. Their recordings of thermodynamic data, made available by NCAR, will be compared for a portion of the first box circuit, from about 1437 GMT, when the aircraft were near the southwest corner D_3, until 1500 GMT, when they reached the northwest corner D_4, the UK-C130 following 25 min behind the other two. They flew northward through the arcs shown

TABLE 3. Radar calibration [from Warner and Austin (1978); column 3 is added from Austin and Geotis (1979)].

Alpha-numeric	Reflectivity (dBZ)	Rainfall rate (mm h^{-1}) from		
		180 $R^{1.45}$	170 $R^{1.48}$	230 $R^{1.55}$
1	11	0.14	0.18	0.098
2	13	0.20	0.24	0.14
3	14	0.23	0.28	0.17
4	16	0.33	0.39	0.25
5	17	0.39	0.45	0.30
6	18.5	0.50	0.56	0.39
7	20	0.65	0.71	0.51
8	22	0.91	0.96	0.74
9	24	1.28	1.29	1.07
A	25.5	1.65	1.62	1.42
B	27.5	2.3	2.2	2.0
C	29	3.0	2.8	2.7
D	30.5	3.9	3.5	3.6
E	33	5.9	5.1	5.6
F	35.5	9.1	7.4	8.9
G	38	13.9	10.8	14.2
H	41	23	17	25
K	44	39	27	43
L	46	55	36	62
M	48	77	49	89

in Figs. 6b and 6c. Time series plots of data from the aircraft, with time running from right to left, are shown in Fig. 11.

Looking in Fig. 11 at temperatures (T) and dew points (T_D) for cases of cloud penetrations, it seems that the aircraft instruments responded well to sudden changes. Effects of rain on temperature measurements—a notorious problem—are not involved here. Penetrated clouds were small; therefore effects of sensor wetting are neglected. At the right of Fig. 11 uncertainty bars for density may be found: an increase in density corresponding to the bars could be caused by changes of −0.2°C in T or −1.2°C in T_D. Changes in plotted densities slightly exceeded those indicated by the uncertainty bars. Potential temperatures fluctuated relatively widely.

During the intervals shown in Fig. 11, the aircraft flew first through the arc shown in Fig. 6c, and then through that shown in Fig. 6b. Using the McIDAS system to superimpose the aircraft track D_3 to D_4 on full resolution visible satellite images at 15 min intervals, the times of passage of the three aircraft through the arcs shown in Figs. 6c and 6b were determined. The arc in Fig. 6c was ~4.5 km wide in the satellite view; cross-hatching in the middle of Fig. 11 indicates its position from the satellite. Cross-hatching near the left of Fig. 11, at later times, likewise shows the passages through the leading edge of the arc shown in Fig. 6b.

Comparing in Fig. 11 the hatched position from the satellite images with the cloud data from the Electra and DC-6 (central strips in the diagram), there is close correspondence between the hatching and the notations C which indicate flight through an environment containing cumulus clouds penetrating through the flight level. This environment is illustrated in Fig. 12, a photograph from the nose camera of the DC-6 at 1448:19 GMT. The notations of cumulus cloudiness (C) show that cumulus convection extended in latitude further than the brightest clouds picked out in the satellite images.

Comparing the notations of cumulus cloudiness and tiny cumulus (C and T in Fig. 11) with the traces of temperature and dew point (T and T_D), it seems that the air of the arcs, containing clouds, featured relatively low temperatures and high dew points at the levels of the DC-6 and the Electra. Near the surface with the UK-C130 such anomalies do not appear at the times of passage of the arcs indicated from the satellite images. Shortly after 1510 GMT the UK-C130 briefly encountered low temperatures and high dew points, with changes of wind. Satellite photographs indicate that probably these phenomena were associated with activity to the south of the arc shown in Fig. 6c.

The arc to the north is relatively difficult to recog-

TABLE 4. Cloud and sounding data from Fig. 11 for arc shown in Fig. 6c (Schematic is shown in Fig. 16). Wind corrections b_u, b_v are from both box circuits. [Lower values are from Grossman's (1977) method.]

Cloud tops		Tiny Cu	Cu med (arc) (~2000 m)	Clear	b_u, b_v (m s^{-1})
Winds at 1067 m	u, v	3.0, 3.5	2.5, 2.8	3.0, 2.0	0.6, 0.3
	m s^{-1} deg^{-1}	4.6/220	3.8/221	3.6/236	(0.5, 0.2)
$-\partial T/\partial z$	(K km^{-1})	5.6	5.8	5.8	
Cloud tops	(m)	~800			
Winds at 537 m	u, v	3.2, 7.6	2.9, 7.3	3.2, 7.0	2.6, −0.1
	m s^{-1} deg^{-1}	8.2/203	7.8/201	7.7/204	(1.7, −0.5)
Cloud bases	(m)	~500	~500		
$-\partial T/\partial z$	(K km^{-1})	~9.0	~9.0	~7.7	
Winds at 238 m	u, v	3.7, 4.5	3.5, 4.2	3.3, 4.8	1.6, −0.1
	m s^{-1} deg^{-1}	5.9/219	5.5/219	5.9/214	(0.5, −0.2)

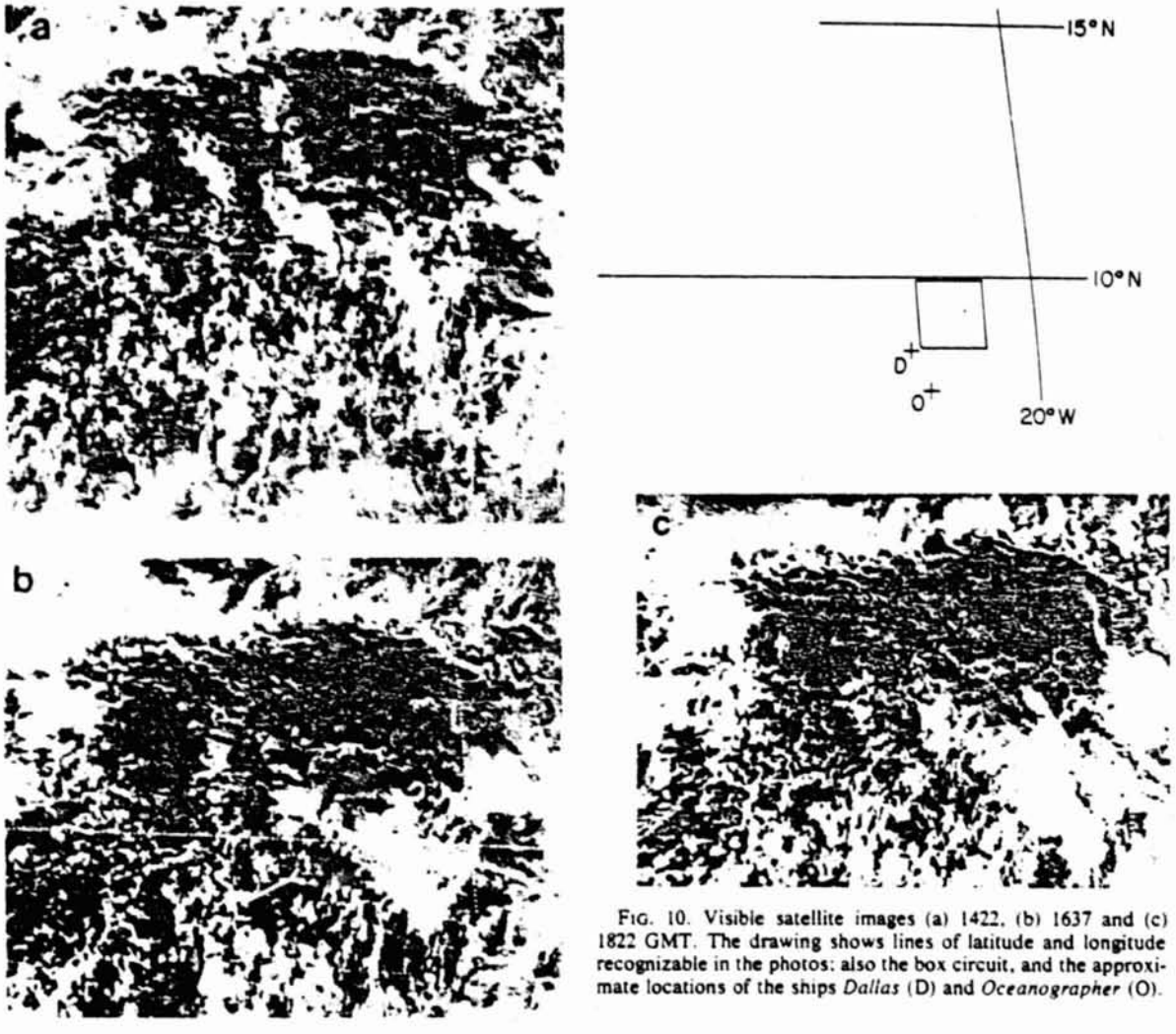


FIG. 10. Visible satellite images (a) 1422, (b) 1637 and (c) 1822 GMT. The drawing shows lines of latitude and longitude recognizable in the photos; also the box circuit, and the approximate locations of the ships *Dallas* (D) and *Oceanographer* (O).

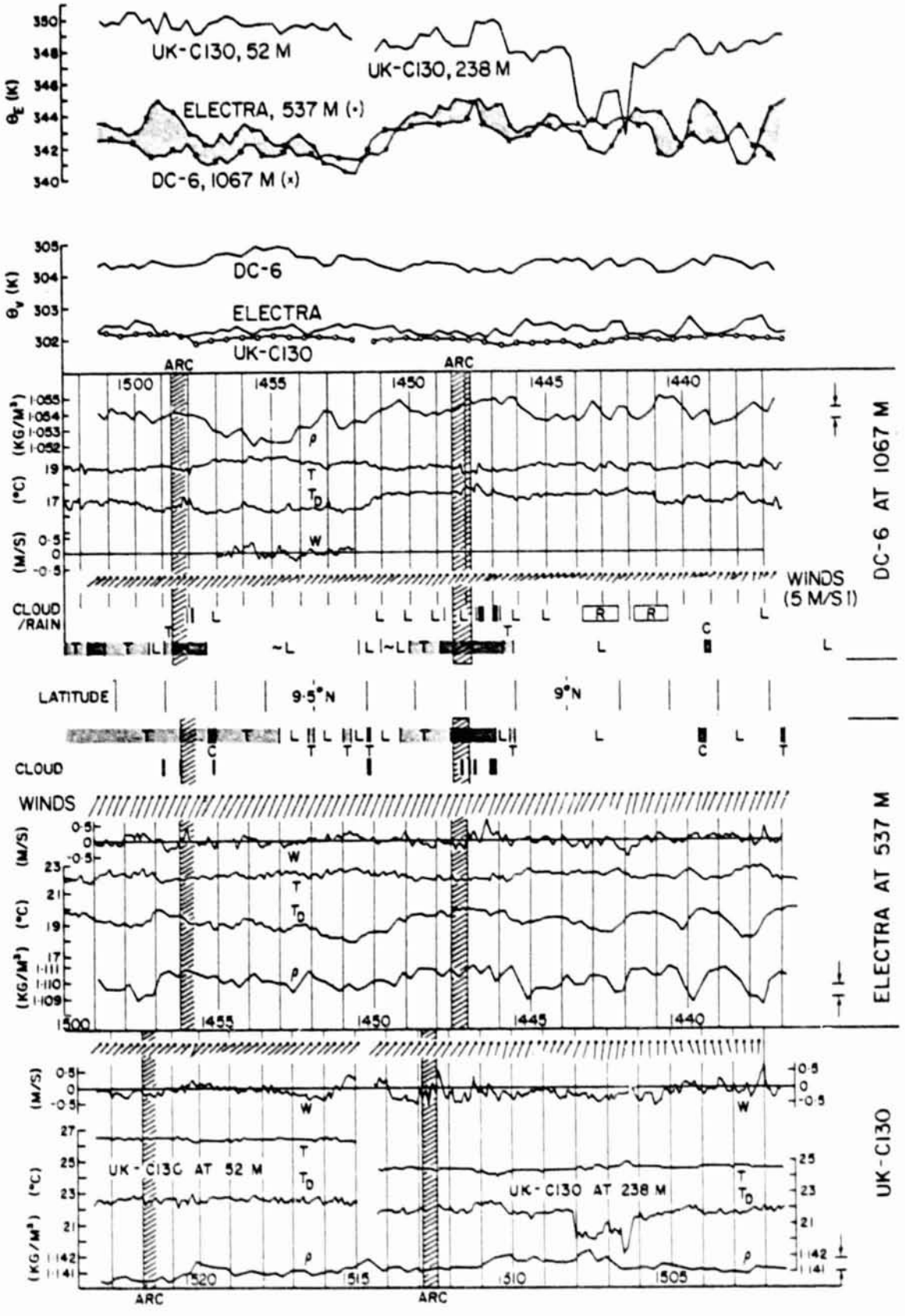
FIG. 11. Time-series plots of data from three aircraft, for the west side of the first box circuit. The plots are keyed to geographical location, shown in the middle. (Note that 0.1° of latitude = 11.1 km.) Cross-hatching indicates the positions of the arcs shown in Figs. 6b and 6c, as derived from full resolution visible satellite images.

At the base is time for the UK-C130. The first plot upward from the base is the density (ρ) of air sampled by the UK-C130 at 15 s intervals, based on 15 s averages (and normalized by transfer to the nominal pressure 985 hPa at constant potential temperature). The next two plots are dew point (T_D) and temperature (T), respectively, plotted every 6 s from 1 s averages (without adjustment for pressure variations). Next come vertical air velocities (W), plotted every 6 s from 6 s averages. Next come winds drawn as short barbs (at 15 s intervals from 15 s averages), after applying the corrections given in Table 4. Winds were from roughly the south: north is upward in this representation of winds; a scale higher up in the diagram on the right shows the length of barb of speed 5 m s^{-1}. The next strip going upward is time for the Electra, then density from the Electra (with normalization to 949 hPa); then T_D, T, W and winds. From the nominal vertical air velocities W from the Electra, a constant 4.6 m s^{-1} has been subtracted. The next strip shows blocks corresponding to times when the Electra was in cloud; above that is a characterization of the convective field through which the aircraft was flying: L means clear air; T (light shading) means that tiny cumulus surrounded the Electra; C (darker shading) means that cumulus mediocris surrounded the aircraft. The central strip in the diagram shows the latitude as the aircraft flew northward at longitude 22.4°W.

The upper middle part of the diagram shows data plotted for the DC-6 in the same manner as for the Electra. The vertical air velocities shown (not available from the NCAR microfilm 1 s data) were abstracted from data over one short flight leg. The mean over the leg may not have been exactly zero, so its removal may mean a (very small) systematic bias.

The DC-6 characterization is slightly different from that of the Electra. The DC-6 flew for much of the time below stratus, and often in mist. In the first strip upward from the middle of the diagram, L means that the air below was clear; T (light shading) means that tiny cumulus were below; C (darker shading) means that cumulus reached 1067 m, flight level. In the second strip, rain (as opposed to the cloud) is indicated by the letter R. Times on the minute when the sky overhead was clear are denoted by L; at other times there was stratus overhead. DC-6 densities were normalized to 891 hPa.

At the top of the figure are plots of virtual potential temperature θ_v and of pseudo-equivalent potential temperature θ_E [calculated by Simpson's (1978) method]. Plots for the different aircraft are superimposed, at 15 s intervals from 15 s averages.


UK-C130, 52 M
UK-C130, 238 M
ELECTRA, 537 M (·)
DC-6, 1067 M (×)
θE (K)
DC-6
ELECTRA
UK-C130
θV (K)
ARC
ARC
1500
1455
1450
1445
1440
(KG/M³)
(°C)
(M/S)
CLOUD /RAIN
WINDS (5 M/S I)
DC-6 AT 1067 M
LATITUDE
9·5°N
9°N
CLOUD
WINDS
ELECTRA AT 537 M
UK-C130 AT 52 M
UK-C130 AT 238 M
UK-C130
1520
1515
1510
1505

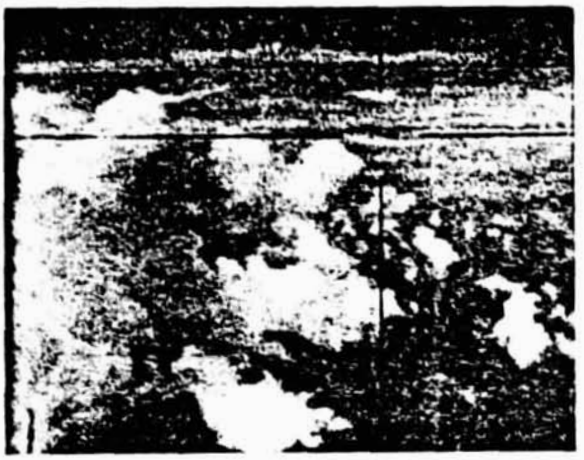


FIG. 12. 1448:19 GMT. Frame from nose camera movie of DC-6 aircraft, taken at altitude about 1067 m while flying northward on the first circuit, within the cumulus environment of the arc shown in Fig. 6c.

nize in Fig. 11. By 1522, when its position was penetrated by the UK-C130, no bright clouds associated with it were present along the aircraft track (see Fig. 6b); it was moving northeastward away from the track and losing its identity.

One other arc was penetrated at low levels, by the Electra flying southward at altitude 35 m at 1612:30 (Fig. 13). As it flew toward the arc, temperatures dropped and dew points rose just as tiny cumulus clouds were encountered. The wind increased and veered (from 6 m s^{-1}/190° to 9.3 m s^{-1}/195°) and then decreased and backed to previous values, within a period of ~1 min or 6 km. This suggests a compact pattern of convergence followed by divergence—of ascent in front with descent (of dense air) behind. The aircraft then ascended from altitude 35 m back up to about 540 m, passing below the base of a vigorous cumulus (at 1614:40), and then entering clear air. This arc was closely associated with precipitating clouds, and the cold gust described may have been associated with precipitation.

Why did the arcs persist for such a long time? First, the traces for the DC-6 and Electra in Fig. 11 show that the environment of clouds of the arcs featured relatively low temperatures and high dew points and air densities. Adjacent clear air featured opposite characteristics, implying that it did not tend to undercut the moist air. Anomalies encountered by the UK-C130 are noteworthy for their absence. (Those at 1507 GMT are discussed below.)

Second, values of θ_v generally increased with height, indicating stability; but it is remarkable that the plots for the UK-C130 and Electra merge in the arcs, indicating neutral stability below cloud base.

Values of θ_E in the arcs decreased with height, indicating convective instability. Weak updrafts (reaching 0.6 m s^{-1}) and downdrafts may be seen in the record from the Electra. On either side of the arcs, above the mixed layer, convective instability decreased to nearly neutral and occasionally stable values. Around the active cumulus of the arcs θ_E was relatively high, probably a result of vertical transport from near the sea. Relatively low absolute values of θ_E on either side of the arc of Fig. 6c, with clear skies (L), high temperatures and low dew points, suggest subsidence.

Third, one may consider air motions in clouds. These generally involve both updrafts and downdrafts on the cloud scale, in adjacent horizontal locations. These cannot readily occur if there is no vertical shear of the wind. With shear, one may visualize updrafts and downdrafts readily coexisting with a circulation having horizontal vorticity (Malkus, 1952). In our case the difference in horizontal wind between the Electra and DC-6 was a few meters per second. Updrafts in the clouds appear to have reached roughly 2 m s^{-1}. It seems that the shear was favorable for the release of instability in small clouds. Fig. 12 shows a close-up view of some of the clouds of the arc of Fig. 6c; individual members of this arc were slender and ephemeral. The clouds appear similar to the small cumuli examined by Telford and Wagner (1974), who found that they moved with a velocity close to that of air near the surface, and featured outward motion near the tops. The clouds seem akin to "hummock" echoes seen by acoustic sounder aboard the *Oceanographer* (Gaynor and Mandics,

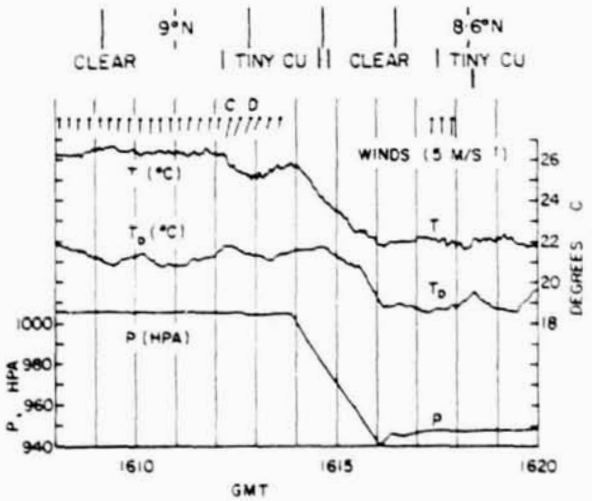


FIG. 13. Time-series plots of data from the Electra, for part of the west side of the second box circuit. At the base is pressure at flight level, then dew points, temperatures and winds plotted in the same way as in Fig. 11. Flight environments of clear air and tiny cumulus are indicated. The Electra flew under the base of a cumulus cloud at about 1614:40 GMT. It penetrated a tiny cumulus at 1618:24. Latitude is shown at the top. Letters C and D next to gusty winds indicate inferred convergence followed by divergence.

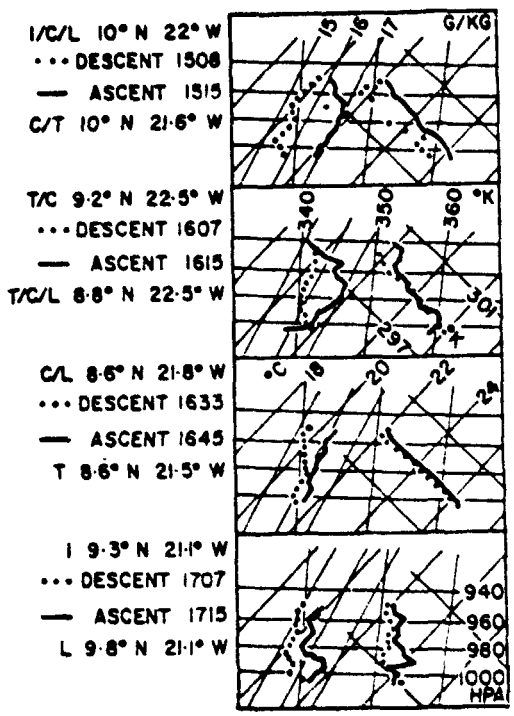


FIG. 14. Soundings from the Electra plotted on Imperial College tephigrams. Isopleths shown of pressure, temperature, specific humidity and dry and pseudo-equivalent potential temperatures. Flight environments, locations and times are indicated (see text for details). See Fig. 4 for locations.

1978). Doppler vertical speed in hummock echoes exceeded 2 m s^{-1}.

Water vapor mixing ratios in the lesser parts of the ranges encountered by each of the three aircraft were compared with rawinsonde and BLIS data (not shown here). These implied that each aircraft encountered air which had subsided from altitudes 500–1000 m above flight level.

Striking in the recordings are the very low dew points and values of θ_E encountered by the UK-C130 at 1507 and 1508, well south of the arc of Fig. 6c. This "dry hole" involved weak descending motion and high densities (although the air was relatively warm). It implies subsidence from a height of several kilometers; apparently a discrete package of air of low θ_E was driven down as a vigorous local return circulation due to a cloudy updraft, and thereafter moved along near the surface. Satellite images show that relatively deep clouds were present near the UK-C130 at this time.

7. Soundings derived from aircraft, and surface conditions

At four times during the two box circuits, the Electra descended from 537 m to 35 m, and returned. Eight soundings were thus obtained, as shown in the tephigrams in Fig. 14. Descents are represented by dots, ascents by continuous lines. The circumstances of the soundings are noted. For instance, "*I/C/L* 10°N 22°W" indicates that the first dotted descent sounding at 1508 GMT was begun in a cloud (I), continued within a field of cumulus mediocris (C), and concluded in clear air (L), near the location 10°N, 22°W (see Fig. 4). The ascent at 1515 started within a field of cumulus mediocris (C) and ended in a field of cumulus (T), roughly 0.4° further east. These soundings indicate that cloudiness was associated with a warm, moist mixed layer, and that clear areas were relatively dry. The last soundings, at 1707 and 1715, were made to the north of the cluster. Its wake was stable.

Conditions in the lowest 10 m above the sea, as assessed from the ship *Dallas*, were broadly undisturbed—meaning an absence of low-lying persistent stable air masses associated with rain. The surface layer did show relatively minor variations associated with small clouds. A singular feature of this day was that the temperatures of the sea, and of the air at 10 m, were alike to within a few tenths of a degree—the measurement uncertainty. Turbulent heat fluxes upward over the lowest 10 m ranged around 70 W m^{-2}, relatively small, and were dominated by the latent rather than sensible heat component.

In order to understand the difference between the regimes of tiny cumulus before the arc shown in Figs. 6c and 11, the arc itself and the clear air behind it, Table 4 is presented. It supplements the data in Fig. 11. The temperature lapse rates ($-\delta T\ \delta z$) in the interval 238–537 m were apparently close to but less than adiabatic with the clouds, and stable in the clear air behind the arc. The next 500 m were stable—similar in all cases.

Winds were corrected for systematic overestimates along (b_x) and across (b_y) the direction of flight (see Grossman, 1977). Careful application of Grossman's method for finding b_x and b_y, by choosing turns in relatively undisturbed air, yielded diverse values. On day 261 the aircraft flew around the box circuit first anticyclonically and then cyclonically, so that measurement errors had an equal and opposite effect in the two cases. This allowed choice of b_x and b_y in such a way as to minimize in a least-squares sense the differences in u and in v between the first and second circuits, for each of the four sides of the box circuit. This method seemed to be the best of several alternatives.

Table 4 shows winds typical of the tiny cumulus, the cumulus and the clear air, at the different levels. Wind corrections b_y and b_x are shown at the right of the table. It may be seen that for the Electra and DC-6, application of Grossman's method in the absence of the information from the circuits would have led to u components 1 m s^{-1} less than those

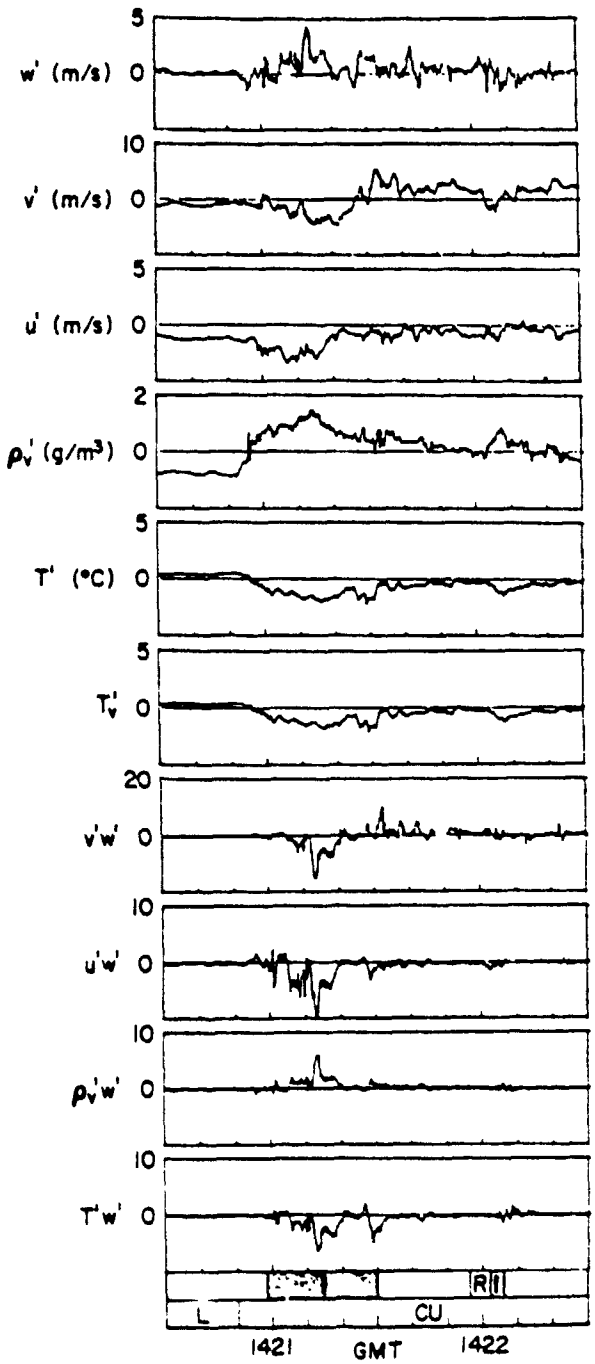


FIG. 15. 1420:30 to 1422:30 GMT. Time-series plots of DC-6 data. From the top: perturbation air speed (m s^{-1}) upward (w'), northward (v') and eastward (u'); water vapor density (g m^{-3}) ρ_v'; temperature T' and virtual temperature T_v' (°C), and fluxes $v'w'$, $u'w'$, $\rho_v'w'$ and $T'w'$. Strips at the bottom show when the DC-6 was in cloud (I) or rain (R), and when it was flying over clear air (L) or among cumulus clouds (CU).

adopted. Such uncertainties limit inferences about the shear of the wind.

It seems safe to conclude that tiny cumulus and cumulus of the arc featured a maximum of wind speed at cloud base (a maximum of v). Perhaps surprising is that wind directions were least at cloud base: a uniform turning with height was expected, but the measurements indicate that this was not the case. That the vertical gradient $\partial u/\partial z < 0$ throughout remains uncertain. The arcs moved from about 198° at 6 m s^{-1} (Section 5). This corresponds with winds near cloud base. Telford and Wagner (1974) found that small cumuli moved at a speed closely related to that of surface air; the present results seem consistent with theirs. Wind speeds decreased in the arcs from front to rear at 537 and 1067 m. This implies vertical confluence into these levels: mesoscale subsidence is inferred, consistent with the speed maximum near cloud base and the stable stratification in the rear of the arcs.

8. Turbulent flux data from aircraft

From the DC-6 aircraft at altitude 1067 m, vertical turbulent fluxes of sensible and latent heat were evaluated. They were means over runs of various lengths, at intervals around both box circuits (Bean *et al.*, 1976). Examining individual traces of perturbation quantities, as in Fig. 15, and comparing these with turbulent fluxes obtained by integration over a total of 35 diverse runs, it was deduced that turbulent fluxes of appreciable magnitude, on the order of hundreds of watts per square meter, were accomplished almost entirely by individual cumulus clouds. All of 18 cloudy updrafts penetrated by the DC-6 during the diverse chosen runs were negatively buoyant. Upward turbulent heat fluxes arose from moist cold updrafts, the moisture being of greater influence than the low temperatures. (Not all cloudy updrafts were measured; those with rain were excluded, owing to deleterious effects on temperature measurements.) The clouds sampled were like those shown in the photograph Fig. 12. They were mostly sampled at an altitude close to their maximum, and were mostly dissipating. Moist cold updrafts near the summits of clouds such as these in the northeast trades were found by Wexler and Malkus (1958). The active stage of a cloud is much briefer than the dissipating stage, so the finding need not be surprising. It is significant here in comparison with findings from the Electra: at 537 m penetrated clouds were roughly neutrally buoyant. The clouds appear to have generated eddy kinetic energy at levels above cloud base but mostly below 1067 m. This is confirmed by weaker updrafts at 1067 m than at 537 m in Fig. 11.

Data from both the DC-6 and the Electra imply that turbulent fluxes in the moist layer are mostly accomplished by cumulus clouds: that mesoscale air circulations were driven primarily by cumulus clouds. The data further lead to the conclusion that the conversion of potential to eddy kinetic energy took place mostly at altitudes between cloud base and 1 km.

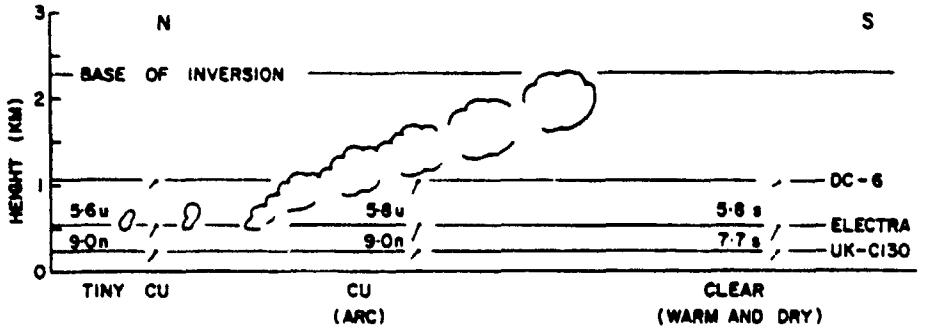


FIG. 16. Schematic diagram of an arc, based on that shown in Figs. 6c and 11. The clouds were composed of a succession of thermals rising in wind shear (see Fig. 5), and have been sketched after the manner of Scorer (1978, pp. 335 and 408). From Table 4, numbers are temperature lapse rates between the levels of flight indicated, and winds are drawn as short barbs as in Fig. 11. From distributions of θ_v and θ_E, the letter S means stable, N neutral, and U convectively unstable. The system moved northward (from right to left). Updrafts occurred at the clouds' leading edges, downdrafts behind. The clear air behind the arc was warm, dry and stably stratified, implying subsidence.

9. Summary and discussion

Conditions of cloudiness in the moist layer, in southerly anticyclonic monsoon flow approaching a growing cloud cluster, have been examined, using airborne and satellite photography, aircraft and ship thermodynamic and wind data, and radar and rawinsonde data.

Unlike middle and upper tropospheric cloud formations, the moist layer featured arc and ring patterns, and rows of small cumulus.

The smallest clouds, called "tiny cumulus", were of linear dimension ~500 m or less; their bases were at 500 to 700 m. These clouds were numerous just in front of arcs of clouds in the moist layer, with separations of 1 or 2 km along and between rows aligned with the wind shear. The data are consistent with a finding of LeMone and Pennell (1976, Fig. 16a) that such clouds resulted from forcing below cloud base associated with roll circulations. Such patterning appeared to occur mostly along the direction of the flow. Also, tiny cumulus were aligned parallel to the arcs, i.e., across the direction of the flow.

The arcs of clouds were very striking. A schematic drawing of one is given in Fig. 16. They were composed of cumulus mediocris reaching roughly the top of the moist layer (2–3 km). Winds near the surface were toward the inside of the arcs. The arcs moved with a velocity close to that of the winds near cloud base. While composed of cumulus mediocris individually short-lived, the patterns persisted for more than an hour traveling steadily. They appeared near radar-detected precipitation. They dissolved finally into patches of cumulus congestus, with radar-detected showers.

Two of the arcs were penetrated by the DC-6 aircraft at 1067 m, the Electra at 537 m and the UK-C130 at 238 and 52 m. At the higher two levels the environment of clouds of the arcs featured low temperatures and high dew points and densities, while clear air on either side of the arcs featured opposite characteristics. There were no anomalies in the mixed layer which could firmly be linked with the arcs shown in Figs. 6b and 6c. One other arc penetrated at altitude 35 m by the Electra showed a convergence-divergence couple involving dense air, of dimension about 6 km; this arc was near radar echoes at the time of sampling.

Virtual potential temperature θ_v increased with height, indicating stability, except between low levels and 537 m in the arcs, where θ_v became about the same at 537 m as in the mixed layer (indicating neutral stability).

Near the clouds, θ_E decreased with height through the aircraft levels, indicating convective instability. In clear air, between 537 and 1067 m, this lapse rate sometimes became stable. The above stability data are tentatively represented by the annotations stable (S), neutral (N) and convectively unstable (U) in Fig. 16.

The arcs in Figs. 6b and 6c featured a wind speed maximum near cloud base (Table 4). Updraft speeds in the clouds appear to have been comparable with the vector difference of horizontal winds through their depth—a condition favorable to overturning. Individual cumulus in the arcs leaned in the direction of the shear—backward with respect to their motion [as shown by Malkus (1952)].

Aircraft measurements of turbulent vertical fluxes showed that they were dominated by cumulus clouds, and that at 1067 m the latent component of the heat flux was generally greater than, and of opposite sign to, the sensible heat component. All the cloud flux data from the DC-6 involved negatively buoyant clouds at 1067 m. (Fluxes were not obtained in large clouds.) The penetrated clouds appeared to be dissipating, near the end of their growing stage, with upward deceleration in the prevailing stable stratification. At 537 m with the

Electra, penetrated clouds were roughly neutrally buoyant. It appears that most of the generation of eddy kinetic energy by clouds of the moist layer occurred below ~1 km.

In view of the longevity of the arcs, an absence of special effects in the records from UK-C130, and the evidence of generation of eddies of kinetic energy by clouds, it seems appropriate to perceive the arcs as self-sustaining entities rather than merely the cloudy manifestations of an original density current.

Schereschewsky (1975, Fig. 1) presents a photograph like some of ours, showing an arc of tilted clouds bordering one side of a circular clear area. Black (1978) shows two arc patterns which he associates with large cumulus clouds. That in his Figs. 9 and 10 may have been similar to the arcs described here. The arc in Black's Fig. 11 is apparently merely a border between a tiny cumulus regime and clear air, a case different from the active arcs of day 261. It is interesting that Black writes: "Sometimes adjacent cumulus cloud arc lines intersect; this produces very strong ascending motions and generates a new cumulonimbus cloud system, which starts the whole process over again." Section 5 provides documentation of a scale-interaction process reminiscent of this.

Lease and Matthews (1978) document a continental cloud arc with coordinated satellite and aircraft measurements. Their observations seem like ours. They interpret them in terms of continued influence of downdraft air following deep convection, rather than in terms of a self-sustaining disturbance. The extensive tropical downdraft documented by Zipser (1969) persisted as a density current until 0710 LT 1 April 1967, but may have yielded a self-sustaining arc of clouds by 1330, judging by the aircraft data and satellite photographs respectively.

Vonder Haar *et al.* (1968) have found many rings of clouds in the tropics, of inside diameter roughly 60 km, containing both shallow and deep convection, and lasting on the order of 10 h. These were usually associated with less recognizably organized clouds. A complete ring was found also in this work (Fig. 8, 1305–1405 GMT). It moved at the same velocity as other moist layer formations. Other rings may be perceived in Fig. 10.

10. Conclusions

The analyses suggest that the moist layer on day 261 was dominated by features of horizontal dimension roughly 40 km and lifetime roughly 2 h. We infer that arc patterns were triggered by dense downdraft air accompanying rainfall. Bearing the pattern of the initial forcing, the arcs persisted. Apparently they were driven by constantly renewed convection, which refreshed the original density current at the surface. Downdrafts of convective scale were shown by vertical air motions measured by the aircraft, and suggested by short-lived variations of horizontal wind near and below cloud base, by vertical changes of wind and by the appearance of the clouds. Downdrafts on the mesoscale behind the arcs were suggested by clear skies, stable stratification and a reduction of wind speed at 537 and 1067 m behind the arc shown in Fig. 6c. The self-perpetuating arcs of the moist layer, with downdrafts on both the cloud and mesoscale, seem similar to the self-perpetuating squall lines of the troposphere, with downdrafts on both the cloud and mesoscale (Zipser, 1977). The moist-layer convection appears to have had a primary role in stirring the layer. Processes driven from the surface appear to have dominated the lowest 500 m; heat transport further upward appears to have been accomplished largely by the arcs—and patterns of deeper clouds. The presence both of horizontal convergence at the surface of 2×10^{-3} s^{-1} (Jalickee and Ropelewski, 1979) and a mixed layer capped by a stable layer (Figs. 2 and 14) may be reconciled in terms of net upward fluxes in clouds and subsidence in mesoscale clear areas.

Triggered by precipitation, the arcs in turn yielded new centers of rain. When the latter appeared, they locally dominated the low levels. Thus there was interaction between convection on the vertical scale of the moist layer with convection on a scale of roughly 6 km. While the arcs of the moist layer were a component of the forcing of showering clouds, probably the number density of the latter was controlled not by the arcs, but by overall conditions of the mid-troposphere.

Acknowledgments. We thank Geoffrey Austin of McGill University for radar data, and Margaret Anne LeMone of NCAR for Electra flux data and helpful comments. R. E. McGavin and R. F. Hartmann, among others of NOAA's Boundary Layer Dynamics Group, contributed to the DC-6 data. Claude Ronne of the Woods Hole Oceanographic Institution helped us with photography. Ron Holle of the National Hurricane and Experimental Meteorology Laboratory assisted in making the cloud maps, with William Snow and Eriberto Varona of the University of Virginia. David R. Fitzjarrald and Gary Barnes assisted us with ship data. Michael Garstang and Mary Morris have given us continual support. We thank Bob Williams at the National Climatic Data Center, Asheville for prompt and cheering provision of data. We thank Colleen Leary for her very thorough and helpful review

We thank Greg Byrd, Nancy Harrison and Kelly Dolan for help with analyses, and Nancy for much photographic work. Drafting was by Thomas Adams and Jim Hamm, and typing by Sandy Smith and Ginty Kelbe.

Most of this analysis was supported by Grant ATM-74-21701-A02 to the University of Virginia, and Grants ATM-7403466 and ATM-78-05951 to the University of Wisconsin, from the Global Atmospheric Research Program, National Science Foundation, and the U.S. GATE Project Office, National Oceanic and Atmospheric Administration.

REFERENCES

Austin, P. M., and S. G. Geotis, 1979: Raindrop sizes and related parameters for GATE. *J. Appl. Meteor.*, **18**, 569–575.

Bean, B. R., R. O. Gilmer, R. F. Hartmann, R. E. McGavin and R. F. Reinking, 1976: Airborne measurement of vertical boundary layer fluxes of water vapor, sensible heat and momentum during GATE. NOAA Tech. Memo. ERL WMPO-36, 83 pp. [U.S. Govt. Printing Office, No. 1977-779-072].

Black, P. G., 1978: Mesoscale cloud patterns revealed by Apollo-Soyuz photographs. *Bull. Amer. Meteor. Soc.*, **59**, 1409–1419.

Chatters, G. C., and C. C Norton, 1977: Wind sets for GATE from SMS image cloud tracking. Space Sci. Eng. Center Rep., University of Wisconsin, Madison, 8 pp.*

Fitzjarrald, D. E., 1978: Horizontal scales of motion in atmospheric free convection observed during the GATE experiment. *J. Appl. Meteor.*, **17**, 213–221.

Garstang, M., G. D. Emmitt, G. Barnes, D. Fitzjarrald, E. Tollerud and J. D. Brown, 1977: The U.S. GATE tethered balloon system: A discussion of the measurements. Cloud populations and their interaction with the boundary layer. Tech. Rep 2, University of Virginia, Charlottesville, 89 pp.**

Gaynor, J. E., and P. A. Mandics, 1978: Analyses of the tropical marine boundary layer during GATE using acoustic sounder data. *Mon. Wea. Rev.*, **106**, 223–232.

Grossman, R. L., 1977: A procedure for the correction of biases in winds measured from aircraft. *J. Appl. Meteor.*, **16**, 654–658.

Jalickee, J. B., and C. F. Ropelewski, 1979: An objective analysis of the boundary-layer thermodynamic structure during GATE. Part I: Method. *Mon. Wea. Rev.*, **107**, 68–76.

Lease, J. C., and D. A. Matthews, 1978: A multi-sensor three-dimensional analysis of a meso-high's development. *Preprints Conf. Cloud Phys. Atmospheric Electricity*, Issaquah, Amer. Meteor. Soc., 483–489.

LeMone, M. A., and W. T. Pennell, 1976: The relationship of the trade wind cumulus distribution to subcloud-layer fluxes and structure. *Mon. Wea. Rev.*, **104**, 524–539.

Malkus, J. S., 1952. The slopes of cumulus clouds in relation to external wind shear. *Quart J Roy Meteor. Soc.*, **78**, 530–542.

Mosher, R. F., 1975: SMS cloud heights. Final Report on NASA Contract NAS5-23296. Space Sci. Eng. Center, University of Wisconsin, Madison, 25 pp.*

Riehl, H., 1954: *Tropical Meteorology*. McGraw-Hill, 392 pp.

Ronne, C., 1959: On a method of cloud measurement from aircraft motion picture films. Unpubl. ms., Ref. No. 59–29, Woods Hole Oceanographic Inst., 13 pp.**

Schereschewsky, P., 1975: Les tourbillons de Bénard dans l'atmosphère; fréquence, diversité, dissymétrie. Cellules ovales isolées. Cellules géantes de brise de mer. Conséquences pour l'environnement. *Comptes Rendus*, **B281**, 205–208.

Scorer, R. S., 1978: *Environmental Aerodynamics*. Ellis Horwood, Chichester, England (see pp. 335 and 408).

Simpson, R. H., 1978: On the computation of equivalent potential temperature. *Mon. Wea. Rev.*, **106**, 124–130.

Snow, J. W., E. C. Varona and J. Simpson, 1975: GATE intercomparisons by cloud imagery. GATE Inform. Bull. No. 8, NCAR, 27–43.

Suchman, D. and D. W. Martin, 1976: Wind sets from SMS images: As assessment of quality for GATE. *J. Appl. Meteor.*, **15**, 1265–1278.

——, D. W. Martin and J. Simpson, 1977: The evolving circulation of an east Atlantic cloud cluster. *Preprints 11th Tech. Conf. Hurricanes and Tropical Meteorology*, Miami Beach, Amer. Meteor. Soc., 333–338.

Telford, J. W., and P. B. Wagner, 1974: The measurement of horizontal air motion near clouds from aircraft. *J. Atmos. Sci.*, **31**, 2066–2080.

Vonder Haar, T., K. Hanson, V. Suomi and U. Shafrir, 1968: Phenomenology of convective ring clouds in the tropics derived from geosynchronous satellite observations. *Proc. Int. Conf. Cloud Physics*, Toronto, WMO, 549–554.

Warner, C., 1978: Photogrammetry from aircraft nose camera movies. *J. Appl. Meteor.*, **17**, 1416–1420.

——, and G. L. Austin, 1978: Statistics of radar echoes on day 261 of GATE. *Mon. Wea. Rev.*, **106**, 983–994.

——, J. Simpson and G. van Helvoirt, 1978: Shallow and deep convection: day 261 of GATE. Cloud populations and their interactions with the boundary layer. Tech. Rep. 4, University of Virginia, Charlottesville, 92 pp.**

——, ——, G. L. Austin, D Suchman and D W. Martin, 1977: Visual, radar and satellite aspects of GATE clouds on day 261. *Preprints 11th Tech. Conf Hurricanes and Tropical Meteorology*, Miami Beach, Amer. Meteor. Soc., 347–354.

——, ——, G. Van Helvoirt, D. W Martin, D. Suchman and G. L. Austin, 1980: Deep convection on day 261 of GATE. *Mon. Wea. Rev.*, **108** (in press).

Wexler, R., and J. Malkus, 1958: Tropical meteorology, observational studies of tropical clouds: Results of 1956 Caribbean aircraft expedition. Unpubl. ms., Ref No. 58–46, Woods Hole Oceanographic Inst., 122 pp.**

Zipser, E. J., 1969: The role of organized unsaturated convective downdrafts in the structure and rapid decay of an equatorial disturbance. *J Appl. Meteor.*, **8**, 799–814.

——, 1977: Mesoscale and convective-scale downdrafts as distinct components of squall-line structure. *Mon. Wea. Rev.*, **105**, 1568–1589.

* Available from the Space Science and Engineering Center, University of Wisconsin, 1225 W Dayton St., Madison 53706.

** Available from the Dept. of Environmental Sciences, University of Virginia, Charlottesville 22903.

Paper 59

Reprinted from MONTHLY WEATHER REVIEW, Vol. 107, No. 3, March 1979
American Meteorological Society
Printed in U. S. A.

Satellite Detection of a Long Curving Cirrus Plume

RAYMOND WEXLER AND WILLIAM C. SKILLMAN
NASA/Goddard Space Flight Center, Greenbelt, MD 20771

9 November 1978

Cirrus plumes have been known to extend long distances ahead of storms. The case of 23 August 1978 is noteworthy because a long cirrus plume, as detected by the GOES E satellite, made a sharp anticyclonic turn and travelled a total distance of 2800 km from the generating thunder-storm, as determined from satellite imagery.

The surface map for 1800 GMT, 23 August 1978 (Fig. 1) shows an open wave over South Dakota with an east–west warm front extending to Lake Huron. Showers and thunderstorm covered a large area from Minnesota across Lake Michigan into Ontario. The 200 mb maps for 0000 GMT 24 August (Fig. 2) showed a ridge centered over the Great Lakes with the wind flow curving anticyclonically to New England and then toward the southwest over the Atlantic due to a sharp trough off the coast. Wind speeds ranged from about 50 to 70 kt. The rawin-

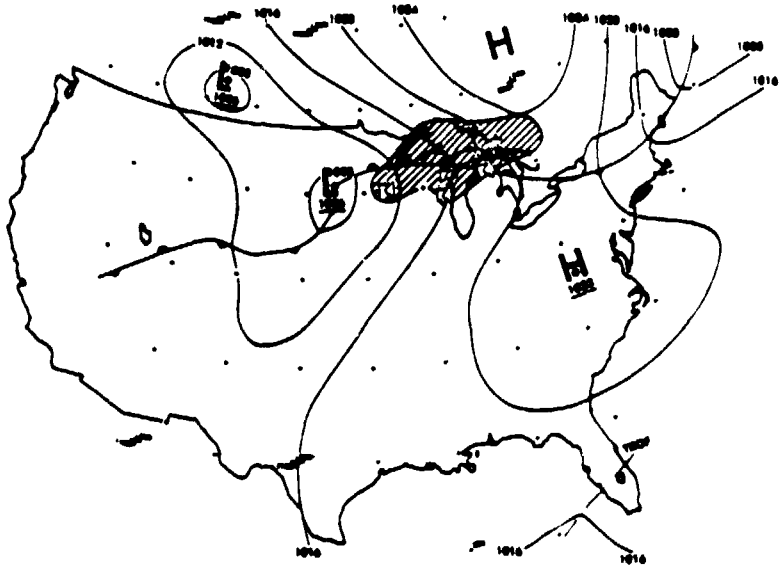


FIG. 1. Surface map 1800 GMT 23 August 1978.

0027-0644/79/030343-04$05.00

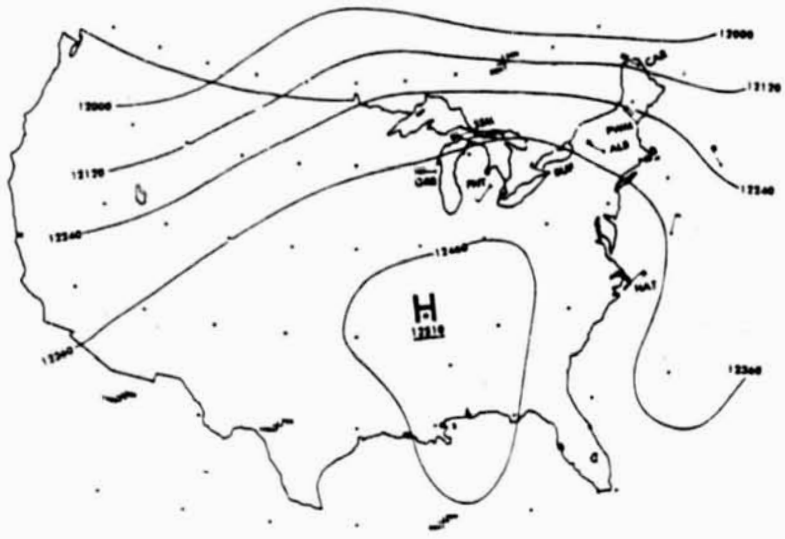
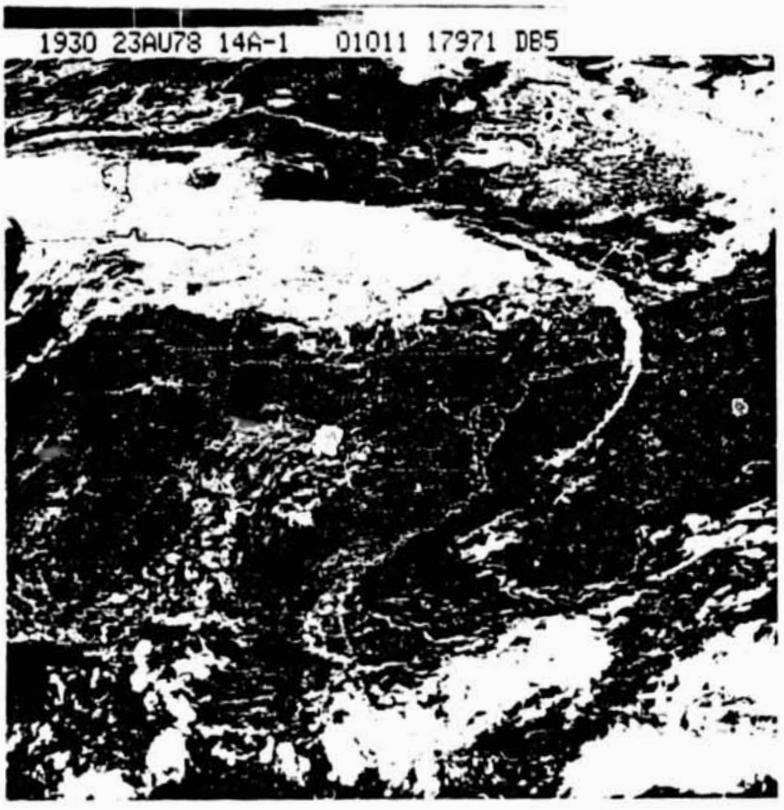


FIG. 2. 200 mb map 0000 GMT 24 August 1978.

sonde at 1200 GMT for Flint, Michigan, in the thunderstorm region shows the tropopause to be at 200 mb.

The GOES E image for 1430 GMT (Fig. 3) showed two cirrus plumes extending from the storm area over the Great Lakes into New York and then curving sharply southward over New England. Five hours later at 1930 GMT (Fig. 4) a single plume extended from the storm area curving toward the southeast over Maine and to the south and southwest off the Atlantic Coast with the tip visible about 300 km ENE of Cape Hatteras. The infrared imagery for 1900 GMT (Fig. 5) showed a similar pattern. The minimum equivalent blackbody temperature (T_B) of the clouds in the active thunderstorms over

FIG. 3. GOES E visible image 1430 GMT 23 August 1978.



Fr . 4. GOES E visible image 1930 GMT 23 August 1978.

FIG. 5. GOES E infrared image 19 . GMT 23 August 1978.

northern Michigan was about 214 K. This temperature corresponds to a height of about 12.5 km (just above the 200 mb level). It may be expected that the cirrus plume would be at about the same altitude. However, because of the lower emissivity of the cirrus plume the observed T_B was considerably higher. In a relatively dense portion of the cirrus off the Massachusetts coast T_B was 226 K. Assuming an emissivity of unity for the thunderstorm cloud top, the emissivity of the dense portion is 0.74.

The speed of movement of the leading edge of a cirrus plume is not necessarily equal to the wind speed because of evaporation of the ice particles. During the 5 h period from 1430 to 1930 GMT (Figs. 3 and 4) the leading edge moved a distance of about 550 km, giving a speed of 30 m s^{-1}. This is in good agreement with the 200 mb winds from New England to the Middle Atlantic states. This may indicate that cirrus evaporation may not have been too important.

The open wave over South Dakota is outlined by a thin band of cloud in the visible imagery at 1930 (Fig. 4). The infrared imagery at 1900 (Fig. 5) shows that these clouds were relatively warm and at relatively low elevations. The cold air to the north of the front was relatively shallow in this region. Further to the east over Michigan the temperature contrast across the front was greater and the cold air deeper and there was considerable overrunning and active convection.

Paper 60

NASA Conference Publication 2076 – Fourth NASA Weather and Climate Program Science Review, January 24-25, 1979, NASA/GSFC, Greenbelt, MD.

Paper No. 16

RAIN OBSERVATIONS IN TROPICAL STORM CORA

T. T. Wilheit, A. T. C. Chang, J. L. King, and E. B. Rodgers, *Goddard Space Flight Center*, R. A. Nieman, B. M. Krupp, and H. Siddalingaish, *Computer Sciences Corporation*, B. C. Diesen, *U.S. Air Force*, and J. Stratigos, *Georgia Institute of Technology*

Passive microwave observations were made in tropical storm Cora at 19.35 and 94GHz. These observations suggest that 94GHz is appropriate for mapping the extent of rain over either land or ocean backgrounds and that some rainfall intensity measurement is also possible.

Paper 61

VOL. 84, NO. C11 JOURNAL OF GEOPHYSICAL RESEARCH NOVEMBER 20, 1979

Trace Gas Concentrations, Intertropical Convergence, Atmospheric Fronts, and Ocean Currents in the Tropical Pacific

P. E. WILKNISS

National Science Foundation, Washington, D C

E. B. RODGERS

NASA, Goddard Space Flight Center, Greenbelt, Maryland

J. W. SWINNERTON, R. E. LARSON, AND R. A. LAMONTAGNE

Naval Research Laboratory, Washington, D C

Shipboard measurements of atmospheric ^{222}Rn, CO, and CH_4 and of dissolved CO in surface waters have been carried out in the equatorial Pacific on a cruise from Ecuador to Hawaii, Tahiti and Panama in March and April of 1974, and during transit from Los Angeles to Antarctica in November and December of 1972. Trace gas results, combined with conventional meteorological data and with satellite images from Nimbus 5 and the defense meteorological satellite project (DMSP), have provided descriptions of the intertropical convergence zones (ITCZ) near 04°N, 102°W and 03°N, 154°W in March of 1974, near 04°N, 86°W in April of 1974, and near 05°N, 139°W in November of 1972. In all cases the ITCZ seems to be located north of the south equatorial current (SEC) as shown by dissolved CO peaks in surface waters In April of 1974 a 'second' ITCZ was observed near 01°S, 102°W just south of the SEC A stationary front near Hawaii (20°N, 147°W) in March of 1974 was investigated. The ITCZ was marked by light shifting winds near a zone of heavy cloud cover and precipitation In the eastern Tropical Pacific atmospheric ^{222}Rn increases distinctly north of the ITCZ and thus serves as an indicator for the ITCZ CO and CH_4 do not always increase coincident with atmospheric ^{222}Rn. The atmospheric features of the stationary front near Hawaii are in many ways similar to those observed for the ITCZ The front is marked by cloud cover, precipitation zone and light shifting winds ^{222}Rn, CO and CH_4 increase significantly behind the front in subsiding air which was traced back to the Asian continent The variation of atmospheric ^{222}Rn, CO and CH_4 with time and geographical area over the equatorial Pacific seems to be a consequence of seasonal variations of the trade wind field and long range transport to the central Pacific from Asia and to the eastern equatorial Pacific from North and Central America

INTRODUCTION

In the tropical Pacific important interactions between the atmosphere and ocean occur, which may greatly influence large-scale weather and climate Because of the importance of the tropical Pacific for the global environment, increased attention has been given recently to investigating the interaction of the tradewind fields with the equatorial current systems [*Barnett*, 1977*a*], for the purpose of forecasting El Niño off the coast of South America because of its economic implications [*Wyrtki*, 1977*a*; *Barnett*, 1977*b*; *Wyrtki et al*., 1976] and to understanding the relationship between atmospheric carbon dioxide and Pacific sea surface temperature as an indicator of air sea interaction processes [*Newell and Weare*, 1977]. Examples of satellite studies are the recognition of long waves at a sea surface temperature front in the tropical Pacific [*Legeckis*, 1977], the global mapping of atmospheric temperature profiles and the distribution of water vapor and liquid water over the oceans [*Staelin et al*., 1977]; the observation of mesoscale eddy dynamics in the eastern tropical Pacific [*Stumpf and Legeckis*, 1977], and the analysis of advection in the Peru Current [*Wyrtki*, 1977*b*] This paper is an attempt to correlate a number of parameters measured from a ship with satellite images. It is not meant to be all inclusive but rather to serve as a stimulant to other researchers in this area. We report here measurements of the trace gases radon (^{222}Rn, a naturally occurring radioactive gas, $t_{1/2}$ = 3.8 days), carbon monoxide (CO, of anthropogenic and natural origin), and methane (CH_4, largely of natural origin) in the atmosphere of the tropical Pacific. Dissolved CO determinations in Pacific surface waters are included. The measurement of CO and CH_4 marine environment background levels is important to understanding the global CO cycle in the biosphere of the oceans and continents [*Linnenbom et al*., 1973, *Seiler and Giehl*, 1977], the impact of anthropogenic CO emissions upon the natural cycle [*Sze*, 1977] and associated changes in the tropospheric trace gas budgets [*Crutzen and Fishman*, 1977], and the possible impact on global climate [*Wang et al*., 1976]

The analytical techniques for the determination of CO, CH_4, and ^{222}Rn are very sensitive, and have good precision and accuracy which allow use of these gases as atmospheric tracers CO can serve well to distinguish between air of marine (0.03–0.10 ppm CO) or continental origin, especially with anthropogenic influence (0.10–1 ppm) [*Wilkniss et al*., 1975] ^{222}Rn is an excellent tracer for continental air over remote ocean areas, because ^{222}Rn originates almost exclusively from the continents, where atmospheric radon levels can reach several hundred pCi m^{-3}, while the ocean is a very weak source of ^{222}Rn, and marine atmospheric background levels may be 1 pCi m^{-3} or less [*Wilkniss et al*., 1974, 1975, *Wilkening and Clements*, 1975]. CH_4 can serve as an indicator of the ITCZ because of its hemispheric differences

Trace gas results of three interhemispheric crossings in March and April 1974 during a cruise from Ecuador to Hawaii, Tahiti and Panama are reported, along with November 1972 results obtained on a cruise from Los Angeles to the Antarctic The 1974 results include analysis of a frontal system

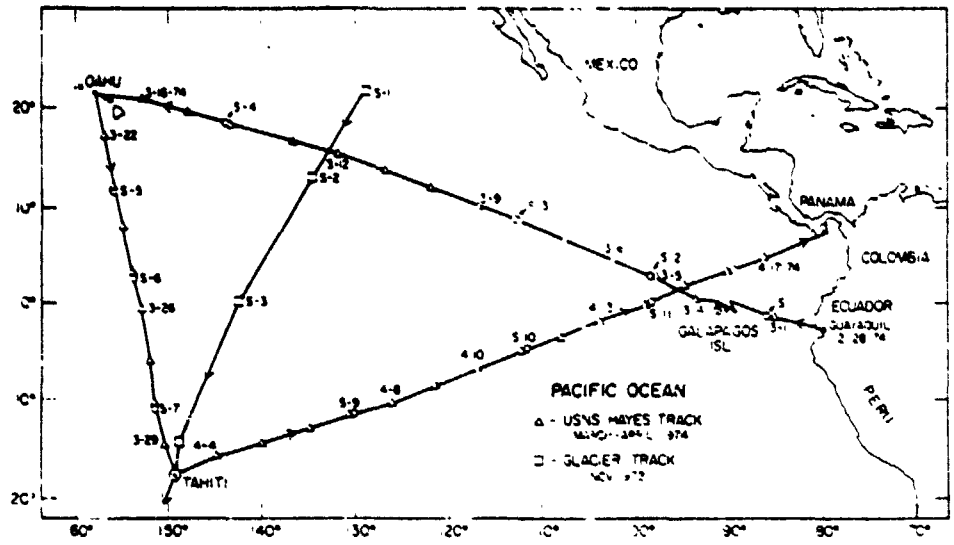


Fig. 1. Cruise track of USNS Hayes in March–April, 1974 and part of cruise track of USCG Glacier in November 1972.

and associated atmospheric trace gas levels near Hawaii. Weather satellite images and conventional meteorological analyses were used to relate trace gas levels to atmospheric phenomena. Dissolved CO measurements and sea surface temperatures were used to delineate equatorial currents.

Experimental

Radon (^{222}Rn) was determined in air samples through radioactivity measurements as described by *Larson* [1973]. CO and CH_4 were determined in air and water samples by gas chromatography according to *Swinnerton et al.* [1968] and *Swinnerton and Linnenbom* [1967]. The cruise track of USNS Hayes in 1972 and part of the USCGS Glacier 1972 track are shown in Figure 1. Meteorological observations taken during the Hayes leg from Guayaquil, Ecuador to Hawaii in March 1974 were analyzed with the aid of Nimbus 5 satellite images and conventional weather maps from the National Meteorological Center. The Nimbus 5 electrically scanning microwave radiometer (ESMR) depicts oceanic precipitation areas [*Wilheit et al.*, 1976; *Kidder and Von der Haar*, 1977] and was used to locate the intertropical convergence zone (ITCZ) and a frontal zone. The Nimbus 5 6.7-μm temperature humidity infrared radiometer (THIR) images were useful in analyzing tropospheric dynamics near a frontal zone [*Rogers et al.*, 1976]. Throughout the 1974 Hayes cruise, daily images (visible and infrared) were obtained from the U.S. Air Force defense meteorological satellite project (DMSP).

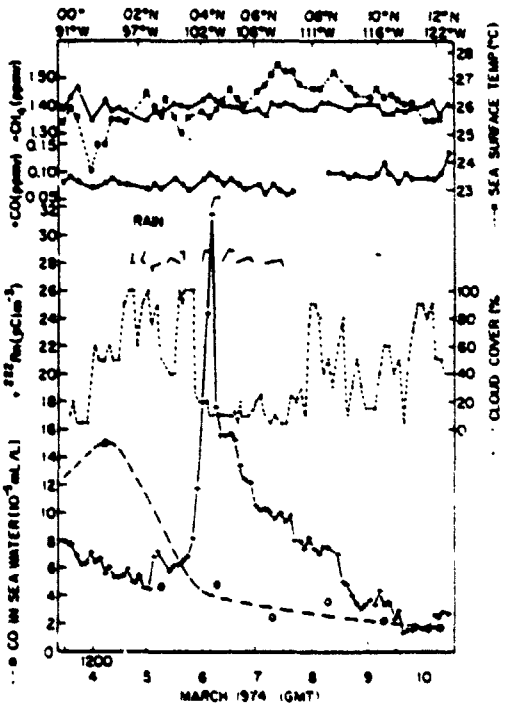


Fig. 2. Trace gas and meteorological data during transit of ITCZ in March 1974

Results and Discussion

Ecuador to Hawaii, March 1974, ITCZ near 04°N, 102°W

A sharp rise to high ^{222}Rn levels [illegible]arch 6, 1200 GMT (Figure 2), implies the arrival of cor[illegible] air, because ^{222}Rn originates almost exclusively from th[illegible]ontinents while the ocean surface is only a weak (1%) source [*Wilkniss et al.*, 1974, 1975; *Wilkening and Clements*, 1975]. No significant changes of atmospheric CO and CH_4 coincide with the ^{222}Rn peak, which suggests a continental source lacking in vegetation and industrial activity. The increase in ^{222}Rn at 04°N, accompanied by a local wind shift from S and W to NW on March 5 (GMT) and to N and finally persistently NE on the morning of March 6 (GMT) indicates the location of the ITCZ at this latitude. The DMSP visible image taken at 1557 GMT on March 6 shows the northern edge of the maximum cloudiness zone (MCZ) [*Sadler et al.*, 1976] at about 04°N, consistent with cloud cover observations aboard Hayes (Figure 2). Two ESMR images taken early on March 7 (Figure 3) clearly depict the location of the ITCZ by a band of precipitation. Radon increased after the rain stopped on March 6. The location of Hayes south and north of the ITCZ and surface trajectories derived from conventional meteorological analyses (Figure 6) have been superimposed on the ESMR image in Figure 3. Figure 6 shows that southeast trade winds from the southern hemisphere, originating near the coast of South America,

dominated the flow south of the ITCZ. North of the ITCZ northeasterly trades originating over southern Mexico account for the elevated ^{222}Rn. CO and CH_4 exhibit similar behavior on both sides of the ITCZ. Dissolved CO levels in surface sea water [*Swinnerton et al.*, 1976] (Figure 2) show a clear maximum, and the peak coincides with the westerly flowing south equatorial current (SEC) as shown by *Swinnerton et al.* [1976]. The coldest surface water temperature was observed at the same location as the dissolved CO peak (Figure 2). The ITCZ is located near the northern boundary of the SEC. This is consistent with earlier observations which relate the ITCZ location to the equatorial current system [*Wyrtki et al.*, 1976; *Wyrtki*, 1974; *Tsuchiya*, 1974]. The location of the ITCZ appears to be between the boundary of the colder SEC and the warmer north equatorial current (NEC) is also apparent from the sea surface temperature. The north equatorial counter current (NECC) is probably not developed at this time of the year when the ITCZ is at its southernmost location [*Wyrtki*, 1974; *Tsuchiya*, 1974].

Tahiti to Panama, April 1974, ITCZ near 04°N; 86°W

Results obtained on the Hayes leg from Tahiti to Panama in April 1974 near the equator are shown in Figure 4. Radon increases distinctly in two steps. On April 13 and 14 the ^{222}Rn rises from South Pacific oceanic background levels to about 6 pCi m^{-3} indicative of continental influence. CO increases to an average of about 0.12 ppm which is more than double the previously prevailing southern hemisphere oceanic background levels. CH_4 also increase significantly from an average of about 1.35 ppm to an average of about 1.43 ppm. The second sharp ^{222}Rn peak on April 17 indicates substantial continental influence but CO and CH_4 remain unchanged. At 0000 GMT on April 14 the local winds measured aboard ship changed from persistently ESE to persistently ENE, indicating a probable area of horizontal convergence near 1°S 102°W. On the afternoon of April 15 and throughout April 16 the local atmosphere was almost calm. Light airs encountered at 0000 GMT on April 17 changed from ESE to ENE and NE, indicating another probable horizontal convergence at 5°N 86°W.

DMSP images on April 5, 6, 7, and from 13 through 17 show two distinct maximum cloudiness zones (MCZ's) [*Sadler et al.*, 1976] at about 5°S and 5°N of the equator with a zone of low cloudiness extending between the MCZ's from the coast of South American to about 110°W along the equator. This is also evident in the 0500 GMT March 7 ESMR image (Figure 3). The cloud band at 5°S has been previously observed in January through March satellite averages [*Anderson et al.*, 1974]. Its strength seems to vary and at times it does not develop at all [*Anderson et al.*, 1974]. The typical features are

Fig. 3. Satellite microwave data for March 7, 1974. Darkness over ocean areas is proportional to liquid water content, and the location of the ITCZ is shown by the line of rain.

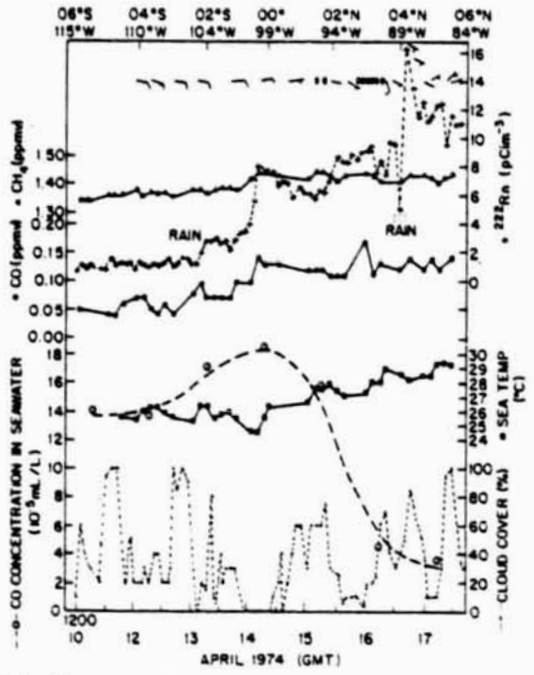


Fig. 4. Trace gas and meteorological data in the equatorial area in April 1974.

shown in the DMSP image (visible) taken on April 17 at 1748 GMT, with the location of Hayes at about 5°N 86°W marked (Figure 5). ^{222}Rn, local wind, cloud cover observations aboard ship, and the DMSP image show the location of the ITCZ between 4° and 5°N at 86°W a. the northern edge of the MCZ in this area. ^{222}Rn, CO, and CH_4 increases on April 13 and 14 occur at the northern edge of the southern MCZ at about 2°S 104°W (Figure 5) just before entering the low cloudiness zone which is verified by ship observations (Figure 4). The ITCZ north of the equator at 4–5°N, 86°W is located north of the SEC as indicated by the dissolved CO values in Figure 4 [*Swinnerton et al.*, 1976] and also found in the March 1974 data (Figure 2). The CO peak in the SEC is better documented in the April data and coincides with the coldest sea surface temperatures and cloud free skies (Figures 4 and 5). The first ^{222}Rn rise and the atmospheric CO and CH_4 increases occur at the southern dissolved CO peak flank, where sea surface temperatures decrease to their lowest observed levels probably indicating the most pronounced portion of the SEC and input from the Peru Current [*Wyrtki*, 1965]. It is tempting to attribute the significant atmospheric CO increase to emanation of the gas from the SEC with its greatly supersaturated CO levels. It must be cautioned, however, that such oceanic supersaturation is not necessarily accompanied by enhancement of CO in the overlying air. For example, Figure 2 shows also a substantial peak of dissolved CO in the SEC with no apparent effect on the atmospheric CO level. While a contribution of CO from the ocean surface to the atmosphere cannot be excluded [*Linnenbom et al.*, 1973], it seems more likely the CO is mostly advected with the ^{222}Rn and CH_4 in this area. Atmo-

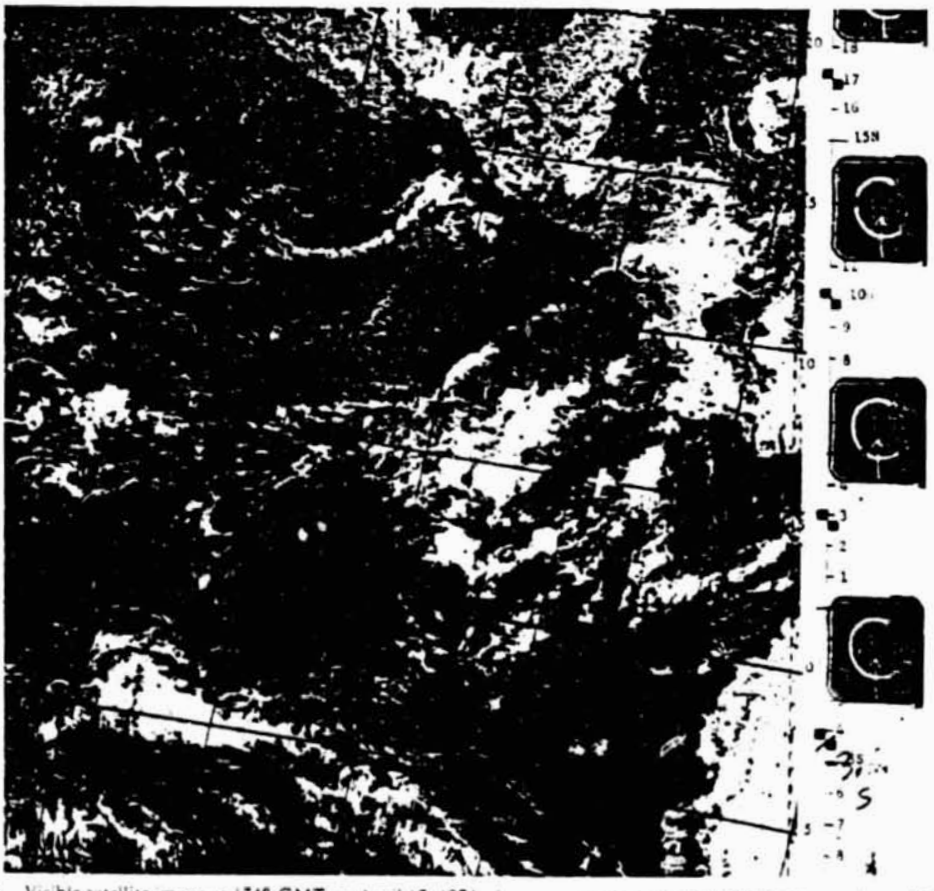


Fig. 5. Visible satellite image at 1748 GMT on April 17, 1974, showing the maximum cloudiness zones at about 5°S and 5°N of the equator. The cross marks the location of the USNS Hayes at 5°N, 86°W.

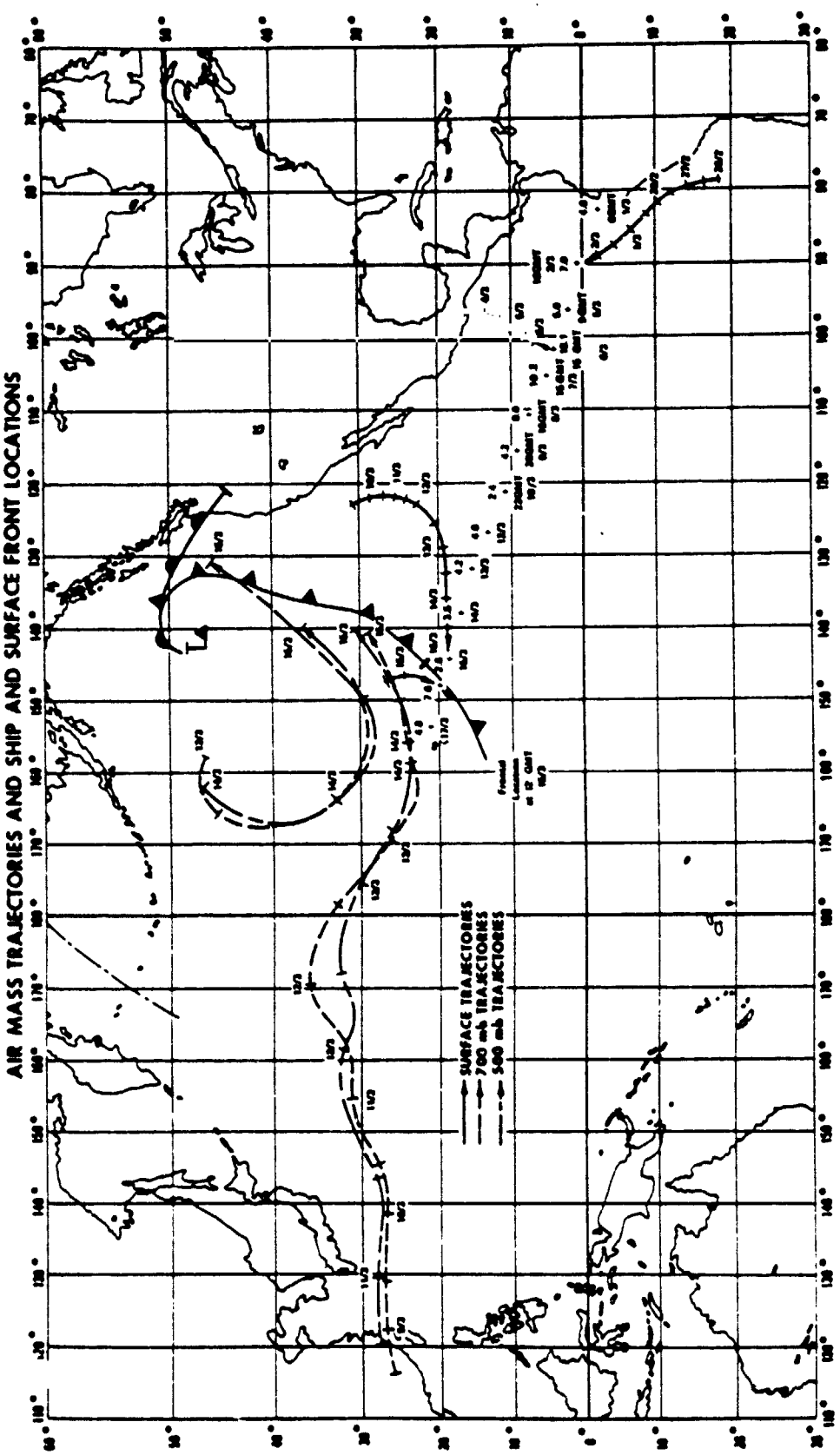


Fig. 6. Ship positions and some air mass trajectories for first leg of cruise in March 1974, shown with frontal location at 1200 GMT March 3, 1974.

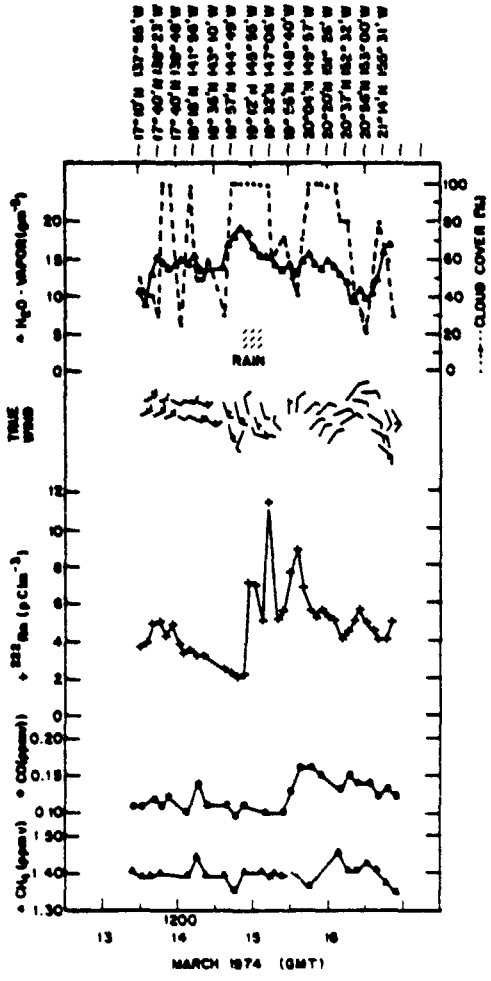


Fig. 7. Atmospheric trace gas and meteorological data during transit of frontal area southeast of Hawaii in March 1974.

spheric ^{222}Rn and CH_4 increases have to be attributed to South America as a source. In comparing the March and April penetration of the ITCZ in this area, it is apparent that trace gas levels are about the same between the southerly and northerly MCZ. Higher ^{222}Rn levels in March 1974 just north of the ITCZ indicate strong flow of air in the NE trades from southern Mexico. The air encountered in April 1974 most likely originated from the coastal areas of South and Central America in the vicinity of the ITCZ. Rain was observed aboard Hayes on April 13, 1974, near the ITCZ at 0900 GMT after which ^{222}Rn, CO and CH_4 increased in the low cloudiness zone. The next rain occurred on April 17 at 0300 GMT after which ^{222}Rn reached the highest peak. Similarly, ^{222}Rn had increased after passing through the precipitation zone on March 5, 1974.

Atmospheric Front, March 1974, near Hawaii at 20°N, 147°W

As the ship proceeded on course toward Hawaii from Central America, it came more and more under the influence of marine air. At 0000 GMT March 15, 1974, as seen in Figure 6, surface trajectories suggest that air is of marine origin which results in low atmospheric radon readings of 2.6 pCi m^{-3}. However, upon crossing the stationary front at about 20°N, the ship moved into a new atmospheric regime. Surface winds aboard ship shifted from SE to NE and the atmospheric Radon increased to 7.6 pCi m^{-3} (Figure 7). Surface trajectories at 0000 GMT March 16, 1974, indicate that air reaching the ship after crossing the front originated from an area of subsiding air of -1 to -2 cm s^{-1} to the north and behind the front as depicted from the NMC vertical velocity charts for 12 GMT March 16, 1974. Geostrophically constructed 700 and 500 mbar trajectories suggest that air subsiding behind the front at 25°N latitude originated from the Asian continent.

Images of this frontal area at 1100 GMT March 15, 1974, from the ESMR and the 6.7 μm THIR sensor on board Nimbus 5 (Figure 8), verify subsidence behind the front depicted by the NMC vertical velocity chart. Since subsiding air brings dryer air downward from aloft, the tropospheric water vapor content decreases in agreement with shipboard observations (Figure 8). The ESMR and the 6.7 μm THIR sensor are sensitive to atmospheric water vapor, where the ESMR sensor observes the integrated atmospheric water vapor as well as liquid water within the lower troposphere below the freezing level [*Wilheit et al.*, 1976], and the 6.7 μm THIR sensor observes the water vapor in cloud-free regions within the middle and upper troposphere at the 600–100 mbar level [*Rodgers et al.*, 1976]. Therefore, these sensors can be used to identify the regions of subsidence and consequent drying. As can be seen in Figure 8, the very bright areas behind the front in the ESMR image depict low values of integrated water vapor in the troposphere, while the dark areas behind the front in the THIR image depict low values of middle and upper tropospheric water vapor.

The comparison of shipboard measurements (Figure 7) with the meteorological analysis (Figures 6 and 8) shows good agreement. On the morning of March 15 (GMT), 2 pCi m^{-3} ^{222}Rn represent northern hemisphere marine background levels. At 0900 GMT March 15, ^{222}Rn begins to rise sharply, eventually to 12 pCi m^{-3}, documenting the arrival of continental air, in this case traced to Asia. Radon fluctuates markedly and decreases but remains definitely above marine air background levels until Hayes reaches Hawaii on March 17. CO also shows a significant increase from 0.10 to 0.16 ppm in the subsiding Asian air behind the stationary front. The CO increase follows the ^{222}Rn increase by about 12 hours, most likely reflecting the independent sources of these two trace gases. While ^{222}Rn is freshly entrained in frontal systems over the continent from rock and soil emanations [*Wilkniss et al.*, 1974], CO is not subject to a short half life as is ^{222}Rn and is of predominantly anthropogenic origin in the northern hemisphere and, therefore, originates from quite a different local source and accumulation pattern. CO also fluctuates and decreases but remains higher than the average CO before the stationary front. CH_4 also shows increases, however they are less well defined and do not correlate as well with the frontal system as CO and ^{222}Rn do.

Rain was observed on the ship between 0700 and 1500 GMT on March 15 at about 19°N, 145.5°W, marking the precipitation zone associated with the front. This is in agreement with the ESMR image (Figure 8) in which the precipitation band is delineated by the dark areas. The ^{222}Rn increased to its highest level just after the ship had passed the precipitation zone and entered an area of decreasing cloud cover (Figure 7), consistent with the transport of ^{222}Rn in subsiding air. The same observations pertain to CO, which also peaks in the relatively cloud free zone. The trace gas data, especially ^{222}Rn, and the water vapor mixing ratio, local wind and cloud cover observations show a complex structure in this southern, diffuse part of the frontal area, in agreement with the ESMR and THIR images (Figure 8).

Hawaii to Tahiti, March 1974, ITCZ near 03°N, 154°W; and Los Angeles to Antarctica, Novem[illegible] 1972, ITCZ near 05°N, 139°W

The shipboard measurements near the equator in March 1974 on the Hayes leg from Hawaii to Tahiti are shown in Figure 9. Results for the 1972 November Glacier equator crossing have been partially replotted from *Swinnerton et al.* [1976] and *Wilkniss et al.* [1973] in Figure 10. A DMSP visible image taken on March 28, 1974, over the central Pacific is shown in Figure 11. Figure 9 shows that ^{222}Rn, CO and CH_4 gradually decrease from the northern into the southern hemisphere. These values reflect the low background levels of northern and southern hemisphere marine air south of about 16°N. Local winds shifted from E and ENE to ESE and SE at about 0400 GMT on March 26 indicating the location of the ITCZ at roughly 3°N, 154°W. Heavy rain occurred at 0600 GMT on March 26, and shower activity was encountered at 1000 and 1400, delineating the precipitation zone associated with the ITCZ. This precipitation zone coincides with the only MCZ observed north of the equator (Figures 9 and 11). From the dissolved CO data, indicating the westward flowing SEC, and the sea surface temperatures, the ITCZ at 03°N, 154°W is located north of the SEC. The atmospheric CO values drop to their lowest and steadiest level at about the location of the dissolved CO peak in the water. The results do not show the SEC to be a measurable source of CO in the atmosphere.

Some of the results of the Hayes cruise in the central Pacific in March 1974 are in striking contrast to those obtained in November 1972 aboard Glacier about 1000 km to the east (Figure 1). In November of 1972 the ITCZ was located at 05°N, 139°W as indicated by the local winds which blew steadily from the NE before 2100 GMT on November 22, and then steadily from the SE after 2200 GMT (Figure 10). Periods of calm winds were recorded at 1400 and 2000 GMT, and light airs from SW-S were observed at 1600 and 1800 hours. The MCZ extended from about 11°N to 5°N, the precipitation zone reached from about 10°N to 5°N. After decreasing, ^{222}Rn and CO attained southern hemisphere marine background levels at the ITCZ, as did CH_4. Especially ^{222}Rn was much higher north of the ITCZ than in 1974, and CO was also elevated in atmospheric samples in the northern hemisphere. The dissolved CO peak exhibits a greater maximum in this area [*Swinnerton et al.*, 1976]. The ITCZ is north of the SEC as in all other cases. Again from Figure 10 the SEC has no apparent effect on the atmospheric CO levels which drop to their lowest values at the SEC. In 1972, ^{222}Rn increases to pronounced continental levels just north of the ITCZ, and CH_4 exhibits typical northern and southern hemisphere background levels north and south of the ITCZ.

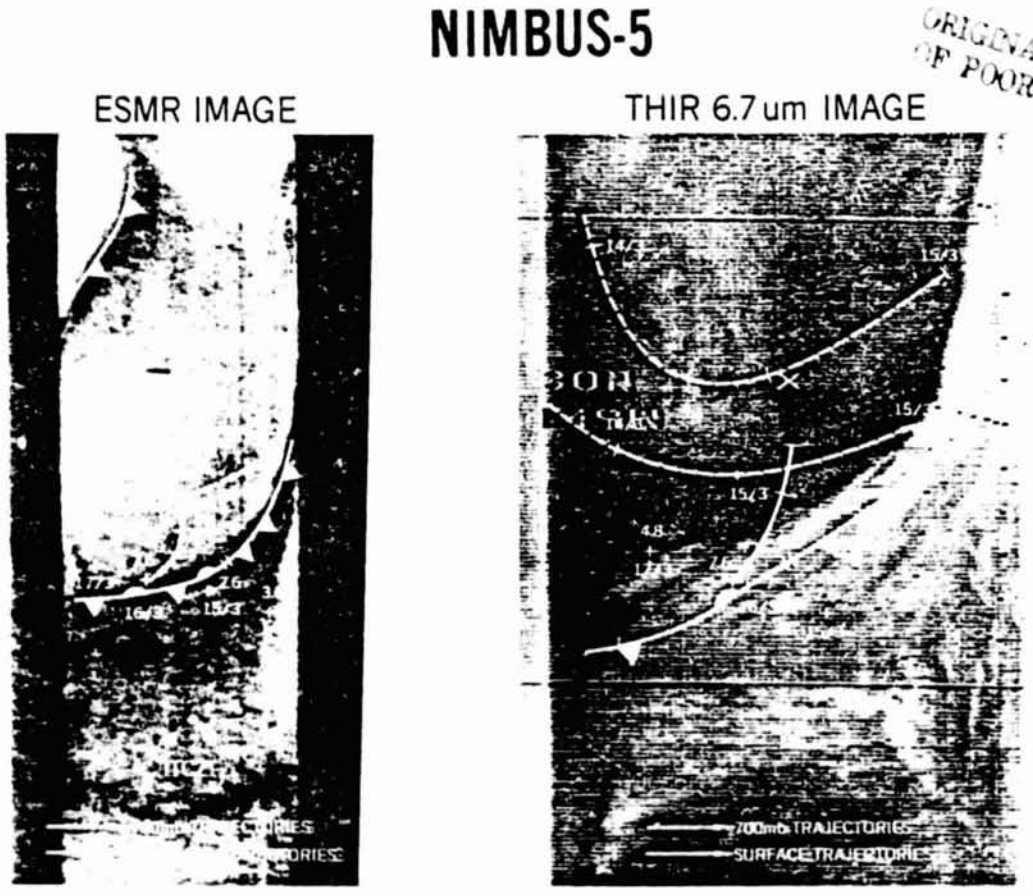


Fig. 8. Nimbus 5 satellite images at 1100 GMT March 15, 1974. Rain associated with ITCZ, cold and occluded fronts are revealed in the ESMR image on the left, while cloud cover and the dry air behind the front are shown in the 6.7 μm image on the left. Frontal lines and air mass trajectories are superimposed on the images.

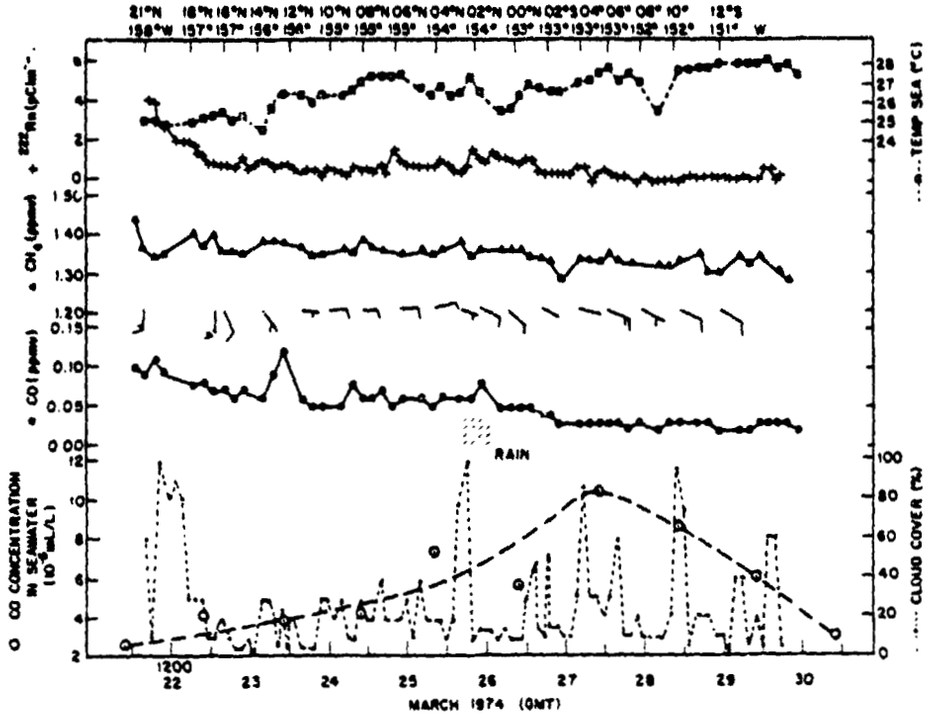


Fig. 9 Trace gas and meteorological data during equatorial crossing from Hawaii to Tahiti in March 1974.

In considering the differences between the 1972 and 1974 results in the central equatorial Pacific, we neglect the 1000 km difference in equator crossings as a major factor in this area, where the east-west extension of the Pacific from Central America to Southeast Asia is about 17,000 km. Rather the observed differences probably result from a combination of several factors. The most likely explanation lies in the seasonal variations of the trade winds systems and the location of the ITCZ [*Wyrtki*, 1965, 1974; *Tsuchiya*, 1974]. In March 1974, 5–10 kn winds from E and ENE gradually changed to 5–10 kn from ESE and SE winds at the ITCZ, no wind minimum was observed. In November 1972, 15–25 kn winds from ENE and NE prevailed from 20°N to 6°N, a wind minimum was observed at 5°N at the ITCZ, while 10–15 kn winds from the SE occurred between 04°N and 02°S.

In the late March 1974 crossing the ITCZ was at a southernmost location at 03°N and not very strongly developed as seen from ship observations and satellite photos (Figures 9 and 11). Therefore, a very gradual change from the northern to southern hemisphere is observed. In November of 1972 the ITCZ was at a more northerly location, and strong NE and SE trade wind fields existed, separated by a wind minimum at the ITCZ. A strong MCZ was observed north of the ITCZ. Thus a better defined atmospheric boundary existed between the northern and southern hemispheres which is reflected and delineated in the atmospheric trace gas levels.

In comparing these results with studies of the trade wind system over the Pacific from long-term observations (1950–1972), the following can be said. According to *Barnett* [1977*a*] and *Wyrtki* [1974] the NE trades are strongest in the first half of the year and weaker in the second half of the year. The southeast trades in the southern hemisphere show a reversed behavior being strong in the second half of the year and weak in the first half. While the present results (Figures 9 and 10) agree with prior observations [*Barnett*, 1977*a*; *Wyrtki*, 1974] as far as the SE trades are concerned, they do not agree for the NE trades, which were much stronger in November of 1972 than in March of 1974. This may be a consequence of short term fluctuations in the trade wind systems or of local disturbances which are not reflected in the 20-year averages of *Barnett* [1977*a*] and *Wyrtki* [1974]. It is interesting to note that 1972 was an 'El Niño' year with November 1972 near the minimum in the southern oscillation index (SOI), while March 1974 measurements occurred near a maximum of the SOI [*Wyrtki*, 1976]. This situation which influences the trade wind system considerably [*Wyrtki*, 1976] may have affected our results.

Summary and Conclusions

Shipboard measurements of atmospheric ^{222}Rn, CO, and CH_4 and of dissolved CO in surface waters have been carried out in the equatorial Pacific. Variations in the ^{222}Rn and CO levels with time and geographical area exceed by far the analytical error limits. ^{222}Rn and CO can serve as tracers for atmospheric phenomena and help to delineate equatorial ocean currents.

Trace gas measurements and standard meteorological observations aboard ship in combination with weather satellite images from Nimbus 5 and the defense meteorological satellite project, and with conventional meteorological charts, make it possible to study small and mesoscale atmospheric phenomena over the Pacific and to relate them to air mass movements.

This combination of methods enable investigations of the ITCZ and of frontal zones over the Pacific. The potential of this approach could be further enhanced by using direct satellite readout aboard ship by 'targeting' measurements and ob-

servations. This has not been the case in the work reported here and, therefore, valuable opportunities to collect pertinent data probably have gone unnoticed.

Some general observations about the ITCZ and the frontal zone near Hawaii include the following:

1. The ITCZ is marked by light, shifting winds near an area of heavy cloud cover (maximum cloudiness zone (MCZ)) and precipitation, and appear to be located north of the south equatorial current (SEC).

2. A 'second' ITCZ with the same atmospheric features was encountered just north of the SEC in the southern hemisphere in April 1974.

3. Atmospheric ^{222}Rn increases north of the ITCZ and serves as a sensitive indicator for this atmospheric boundary. Increases of atmospheric CO and CH_4 do not always accompany those of ^{222}Rn. Observed increases of the atmospheric trace gases occur north of the precipitation zone associated with the ITCZ in areas of clear skies or limited cloud cover, indicating transport in subsiding air.

4. Observations on crossing the stationary front near Hawaii are similar to those made at the ITCZ. The front is marked by light and shifting winds, a heavy cloud cover and precipitation zone, and an adjacent area of greatly reduced cloudiness. ^{222}Rn, CO and CH_4 increase in the atmosphere behind the front and behind the precipitation zone in subsiding air and under reduced cloud cover. A complicated atmospheric structure at the southern, diffuse end of the stationary front is indicated by successive peaks in the atmospheric ^{222}Rn levels. The different source areas of ^{222}Rn and CO on the Asian continent are reflected in the delayed increase of CO compared to ^{222}Rn.

Atmospheric ^{222}Rn, CO and, to a lesser extent, CH_4, vary significantly with time and geographical area over the Pacific north of the ITCZ. This seems to be a consequence of the long-range transport from different source areas where transport and level depend on large-scale atmospheric phenomena which vary seasonally, i.e. the jet stream and cyclonic activity. The transported ^{222}Rn and CO from the Asian continent to the central Pacific north of the ITCZ in late November 1972, compared to late March 1974, and also in the strong features of the ITCZ in 1972 compared to a very weak ITCZ in 1974 illustrate these variations. It is possible that the difference in the southern oscillation index (minimum November 1972, maximum March 1974) has affected conditions in the central Pacific as discussed earlier. In the eastern equatorial Pacific, north of the ITCZ, transport and levels of ^{222}Rn and CO from North and Central America occurs via the NE trades and will vary seasonally with the strength and extent of the trade wind field. The big differences in the ^{222}Rn peaks at 04°N north of the ITCZ in the eastern equatorial Pacific in March and April of 1974 may reflect either enhanced transport in the NE trades from Mexico in March, compared to that from Central America in April or short-term fluctuations such as the possible influence on the NE trades by 'northers,' strong winds from

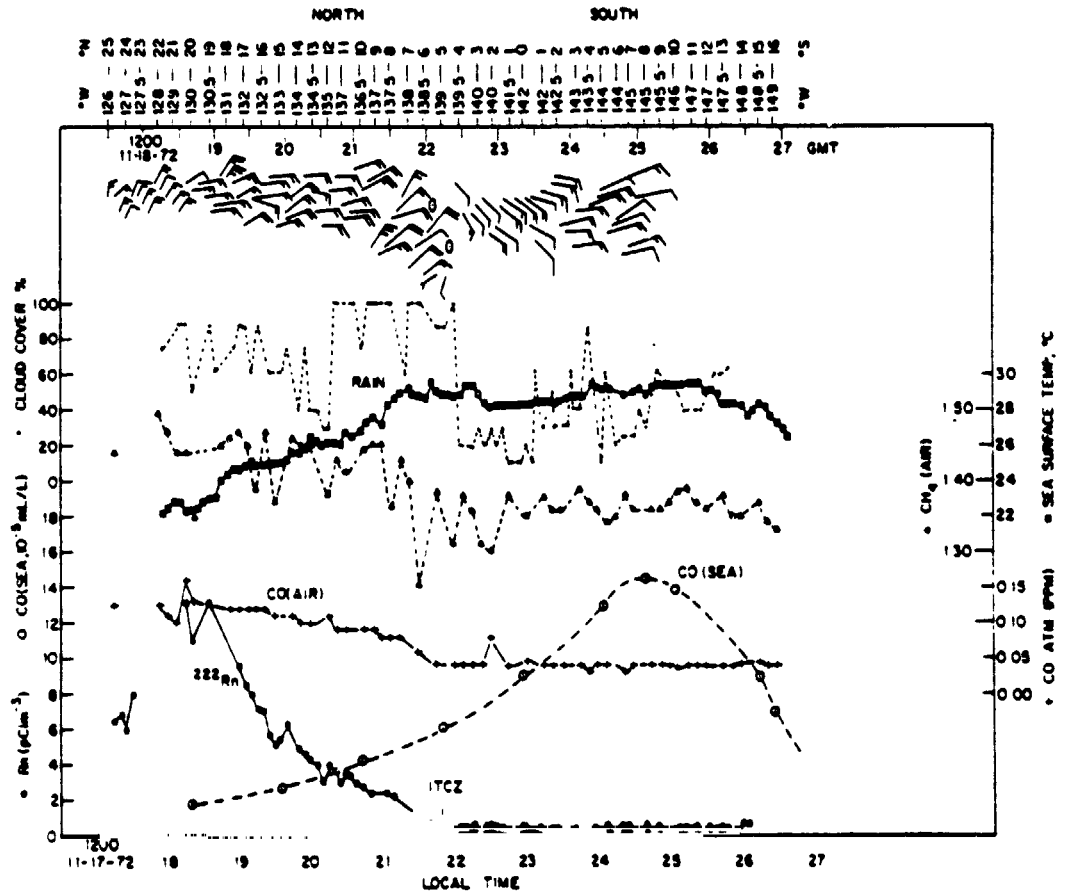


Fig. 10. Trace gas and meteorological data during equatorial crossing in November 1972.

Fig. 11. DMSP visible image at 2109 GMT on March 28, 1974, illustrating the rather cloud free equatorial region throughout the late March transit. The location of the USNS Hayes at 12°S, 151°W is shown by the cross.

the Gulf of Tehuantepec [*Stumpf and Legeckis*, 1977]. These winds are caused by cold air outbreaks driven by anticyclones originating in Canada [*Stumpf and Legeckis*, 1977]. The cold air penetrates south over the Isthmus of Tehuantepec (southern Mexico) and out over the equatorial Pacific. These 'north' wind episodes occur 10–15 times each winter, most frequently from November through April and may be the cause of significant atmospheric trace gas fluctuations over the eastern equatorial Pacific.

Acknowledgments. We thank the U.S. Coast Guard, the Military Sea Life Command, and the masters and crews of USCG Glacier and USNS Hayes, the U.S. Air Force for Defense Meteorological Satellite Project images, and C. H. Cheek, Naval Research Laboratory, and R. C. Taylor, National Science Foundation, for critical reading of the manuscript and valuable discussions.

References

Anderson, R. K., et al. Application of meteorological satellite data in analysis and forecasting, *ESSA Tech. Rep. NESC 51*, p. 4-B-4 Nat. Oceanic and Atmos. Admin., Washington, D. C., March 1974.

Barnett, T. P., The principal time and space scales of the Pacific trade wind fields, *J. Atmos. Sci., 34*, 221–236, 1977*a*.

Barnett, T. P., An attempt to verify some theories of El Niño, *J. Phys. Oceanogr., 7*, 633–647, 1977*b*.

Crutzen, P. J., and J. Fishman, Average concentrations of OH in the tropsphere and the budgets of CH_4, CO, H_2, CH_3CCl_3, *Geophys. Res. Lett., 4*, 321–324, 1977.

Kidder, S. Q., and T. H. Vonder Haar, Seasonal oceanic precipitation frequencies from Nimbus 5 microwave data, *J. Geophys. Res., 82*, 2083–2086, 1977.

Larson, R. E., Measurements of radioactive aerosols using thin plastic scintillators, *Nucl. Instrum. Methods, 108*, 467–470, 1973.

Legeckis, R., Long waves in the eastern equatorial Pacific Ocean: A view from a geostationary satellite *Science, 197*, 1179–1181, 1977.

Linnenbom, V. J., J. W. Swinnerton, and R. A. Lamontagne, The ocean as a source for carbon monoxide, *J. Geophys. Res., 78*, 5333–5340, 1973.

Newell, R. E. and B. C. Weare, A relationship between atmospheric CO_2 and Pacific sea surface temperatures, *Geophys. Res. Lett., 4*, 1–2, 1977.

Rodgers, E. B., V. V. Salomonson, and H. L. Kyle, Upper tropospheric dynamics as reflected in Nimbus 4 THIR 6.7-μm data, *J. Geophys. Res., 81*, 5749–5758, 1976.

Sadler, J. C., L. Oda, and B. J. Kilonsky, Pacific Ocean cloudiness from satellite observations, *UHMET 76-01*, Dep. of Meteorol., Univ. of Hawaii, Oct. 1976.

Seiler, W., and H. Giehl, Influence of plants on the atmospheric carbon monoxide, *Geophys. Res. Lett., 4*, 329–332, 1977.

Staelin, D. H., P. W. Rosenkranz, F. T. Barath, E. J. Johnston, and J. W. Waters, Microwave spectroscopic imagery of the earth, *Science, 197*, 991–993, 1977.

Stumpf, H. G., and R. V. Legeckis, Satellite observation of mesoscale eddy dynamics in the eastern tropical Pacific Ocean, *J. Phys. Oceanogr., 7*, 648–658, 1977.

Swinnerton, J. W., and V. J. Linnenbom, Determination of the C_1-C_4 hydrocarbons in seawater by gas chromatography, *J. Gas Chromatogr., 5*, 570–573, 1967.

Swinnerton, J. W., V. J. Linnenbom, and C. H. Cheek, A sensitive gas chromatographic method for determining CO in seawater, *Limnol. Oceanogr., 13*, 193–196, 1968.

Swinnerton, J. W., R. A. Lamontagne, and W. D. Smith, Carbon monoxide concentrations in surface waters of the east tropical Pacific in 1974: Comparison with values, *Marine Chem., 4*, 57–65, 1976.

Sze, N. D., Anthropogenic CO emissions: Implications for the atmospheric CO-OH-CH_4 cycle, *Science, 195*, 673–675, 1977.

Tsuchiya, M., Variation of the surface geostrophic flow in the eastern intertropical Pacific Ocean, *Fish. Bull.*, *72*, 1075–1086, 1974.
Wang, W. C., Y. L. Yung, A. A. Lacis, T. Mo, and J. E. Hansen, Greenhouse effects due to man-made perturbations of trace gases, *Science*, *194*, 685–690, 1976.
Wilhelt, T. T., J. S. Theon, W. E. Shenk, L. J. Allison, and E. B. Rodgers, Meteorological interpretations of the images from the Nimbus 5 electrically scanned microwave radiometer, *J. Appl. Meteorol.*, *15*, 166–172, 1976.
Wilkening, M. H., and W. H. Clements, Radon 222 from the ocean surface, *J. Geophys. Res.*, *80*, 3828–3830, 1975.
Wilkniss, P. E., R. A. Lamontagne, R. E. Larson, J. W. Swinnerton, C. R. Dickson, and T. Thomas, Atmospheric trace gases in the southern hemisphere, *Nature*, *245*, 45–47, 1973.
Wilkniss, P. E., D. J. Bressan, R. A. Carr, and R. E. Larson, Chemistry of marine aerosols and meteorological influences, *J. Rech. Atmos.*, *8*, 883–893, 1974.
Wilkniss, P. E., D. J. Bressan, R. A. Lamontagne, R. E. Larson, and J. W. Swinnerton, Observation of frontal zones and the measurement of atmospheric trace constituents in coastal areas, paper presented at 3rd Symposium on Meteorological Observations and Instrumentation, Amer. Meteorol. Soc., Washington, D. C., Feb. 1975.
Wyrtki, K., Surface currents of the eastern tropical Pacific Ocean, *Bull. IX(5)*, Inter-Amer. Trop. Tuna Comm., La Jolla, Calif., 1965.
Wyrtki, K., Equatorial currents in the Pacific 1950 to 1970 and their relations to the winds, *J. Phys. Oceanogr.*, *4*, 372–380, 1974.
Wyrtki, K., Sea level during the 1972 El Niño, *J. Phys. Oceanogr.*, *7*, 779–787, 1977*a*.
Wyrtki, K., Advection in the Peru Current as observed by satellite, *J. Geophys. Res.*, *82*, 3939–3943, 1977*b*.
Wyrtki, K., E. Stroup, W. Patzert, R. Williams, and W. Quinn, Predicting and observing El Niño, *Science*, *191*, 343–346, 1976.

(Received May 15, 1978;
revised September 20, 1978,
accepted September 25, 1978.)