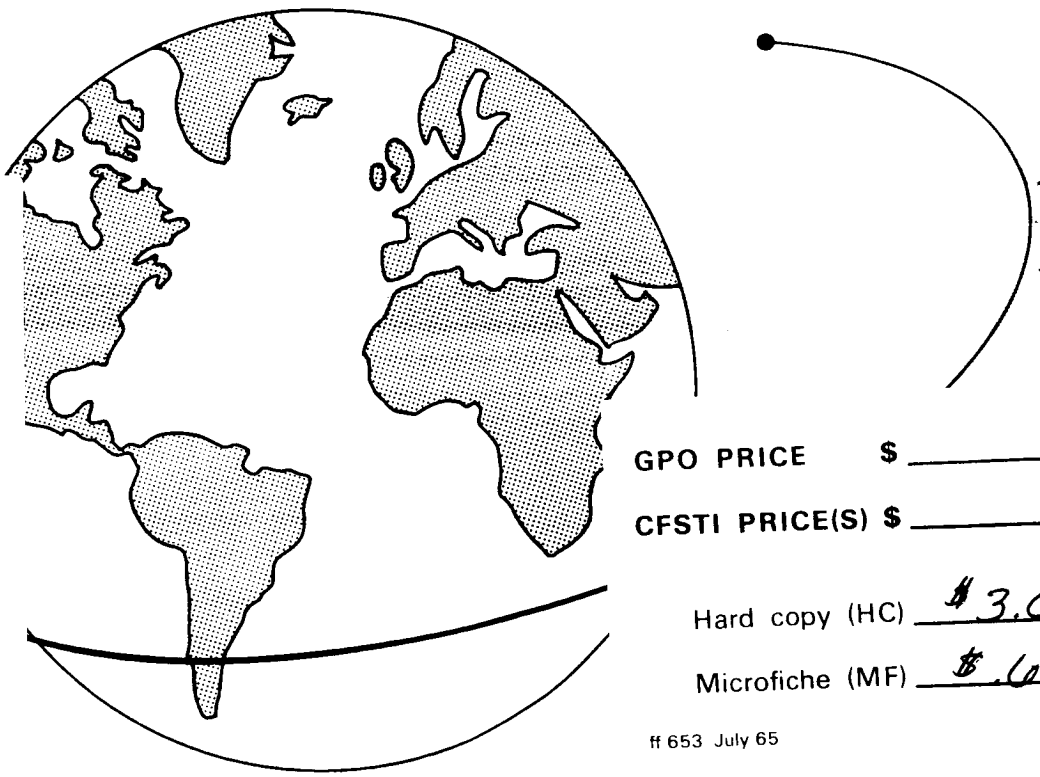
SCIENTIFIC HORIZONS FROM SATELLITE TRACKING

Edited by C.A. LUNDQUIST and H.D. FRIEDMAN



Smithsonian Astrophysical Observatory
SPECIAL REPORT 236

Research in Space Science

SAO Special Report No. 236

SCIENTIFIC HORIZONS FROM SATELLITE TRACKING

edited by

Charles A. Lundquist and Henry D. Friedman

December 30, 1966

Smithsonian Institution
Astrophysical Observatory
Cambridge, Massachusetts, 02138

611-22

The work reported in these papers has been supported in part by Grants NsG 87-60 and NSR 09-015-018 from the National Aeronautics and Space Administration.

TABLE OF CONTENTS

EDITORS' PREFACE

During the summer of 1966, a seminar convened weekly at the Smithsonian Astrophysical Observatory to review and explore the research opportunities offered by satellite tracking. The seminar was essentially an "in-house" function, but visitors were always welcome and several scientists from outside SAO attended more or less regularly. Generally a meeting consisted of a lecture and discussion period totaling 1 1/2 hours, after which many attendees found it convenient to continue their discussions during informal luncheons.

A review of scientific horizons from satellite tracking was timely for several reasons. First, important milestones, such as the Smithsonian Standard Earth, had recently been achieved. Second, newly developed tracking techniques offered greater observational accuracy. Third, SAO had selected a new digital computer for delivery in autumn 1966; consequently, most programs would have to be rewritten, which offered a convenient chance to broaden capabilities and flexibility. Fourth, expanding national and international programs in related fields were generating new interfaces with the SAO scientific efforts.

Each speaker with an SAO affiliation was asked to prepare a manuscript summarizing the topic of his presentation. Intentionally, the manuscripts were not collected until the end of the summer so that the authors would have the full benefit of related lectures and discussions. Various authors elected to prepare papers of differing length and detail, befitting the very different circumstances associated with the diverse topics covered. In the same vein, we have exercised considerable freedom in adding editorial notes, often derived from our introductory remarks or from our interpretation of the discussions.

Another factor influencing the nature of the lectures was the composition of the audience. The seminar included a number of individuals associated with the mechanics and hardware of data collection and processing. This fact probably had more effect on the oral discussions than on the written material.

Much of the understanding that emerged from the seminar has already been important in planning for future programs at SAO. Two recent descriptions of SAO programs (Whipple and Lundquist, 1966; Lundquist, 1966) have drawn substantially upon points established during the seminar.

In closing, we express our thanks to the authors and to members of the SAO Editorial and Publications staff who made possible this collection of manuscripts. Our thanks go also to Dr. S. K. Runcorn, of Kings College, Newcastle-upon-Tyne, who addressed the seminar, and to the many visitors who contributed through their remarks and discussions.

December 1966

Charles A. Lundquist
Henry D. Friedmann

REFERENCES

LUNDQUIST, C. A.
1966. Results from photographic and laser tracking systems. In Proceedings XVIIth International Astronautical Federation Congress, Madrid, in press.

WHIPPLE, F. L., AND LUNDQUIST, C. A.
1966. Tracking by the Smithsonian Astrophysical Observatory. In Proceedings of Discussion Meeting on Orbital Analysis, October, Proc. Roy. Soc. London, in press.

SCIENTIFIC HORIZONS FROM TRACKING SPACE OBJECTS

Charles A. Lundquist[1]

Ten years ago, during the initial planning for the Baker-Nunn camera network, Dr. Fred L. Whipple formulated the scientific objectives of optical satellite tracking (see, for example, Whipple and Hynek, 1958). As initially stated, the objectives were:

> to tie together the observing stations and the center of the geoid to a precision of the order of 10 m, ... to add appreciably to knowledge of the density distribution in the earth, particularly in the crustal volumes, ... to provide precise information as to the density of the atmosphere ... and periodic effects or predictable cyclic effects that may occur in the earth's high atmosphere .

During the years following the first artificial satellite, steady progress was made toward these goals.

One recent milestone at SAO was the adoption of Standard Earth parameters based on tens of thousands of precisely reduced Baker-Nunn observations (Whipple, 1966; Lundquist and Veis, 1966). The adopted parameters include coordinates for the Baker-Nunn cameras (Gaposchkin, 1966a, b; Köhnlein, 1966; Lundquist, 1966; Veis, 1966a, b). As anticipated by Whipple, these coordinates have an uncertainty of 10 to 20 m. A new set of tesseral harmonic coefficients for the geopotential is another ingredient of the SAO Standard Earth (Gaposchkin, 1966a, b; Köhnlein, 1966). In addition, a set of zonal harmonic coefficients has been derived by Yoshihide Kozai (1964). Until significant quantities of new data are collected and analyzed — an effort requiring many months — the Standard Earth parameters will be used at SAO as a basis for further scientific investigation and for routine satellite ephemerides.

GEOS is a typical satellite for which an orbit has been determined from Baker-Nunn observations using the Standard Earth parameters. Its perigee is high enough so that drag has very little effect on the orbit. After the

---

[1]Assistant Director, Science, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts.

determination of coefficients $C_{n,12}$ and $S_{n,12}$, which have "resonant" effects for the GEOS orbit, the deviation between observed and computed station-satellite directions is about 4 arcsec. This is only slightly greater than the estimated 2- to 3-arcsec accuracy for the Baker-Nunn camera and its associated clock.

The information about the high atmosphere derived from satellite-tracking data has surpassed original expectations (Jacchia, 1963, 1966). Dr. L. Jacchia and his associates have produced an atmosphere model that incorporates detailed variations of atmospheric density as a function of several indices of solar and geophysical phenomena (Jacchia, 1965). Most recently, this model has been slightly revised, recognizing that the diurnal atmospheric bulge is apparently elongated in the north-south direction and that its center does not move much from the equator (Jacchia and Slowey, 1966).

Hence, the objectives established 10 years ago by Whipple have been achieved and most have been surpassed. It is also clear that the string has not yet run out, and that Baker-Nunn data and radio data of comparable accuracy can provide the basis for additional investigations. Now is an opportune time to reassess these possibilities and plan the operation of the network to accommodate them. This is one of the purposes of this seminar.

More urgent, however, is the need to establish new objectives corresponding to the greater accuracy provided by recently developed tracking techniques — particularly lasers. This is the primary goal of this seminar.

Graphically, the goal is the construction of a figure giving accuracy of tracking and analysis on one axis and corresponding research objectives on the other (see Figure 1). The Baker-Nunn network performance is indicated by a vertical bar at the 10-m mark. Anticipating a later paper, laser systems currently offer an accuracy of about 1 m and advanced systems may yield some tens of centimeters. These accuracies are likewise indicated by vertical bars in Figure 1.

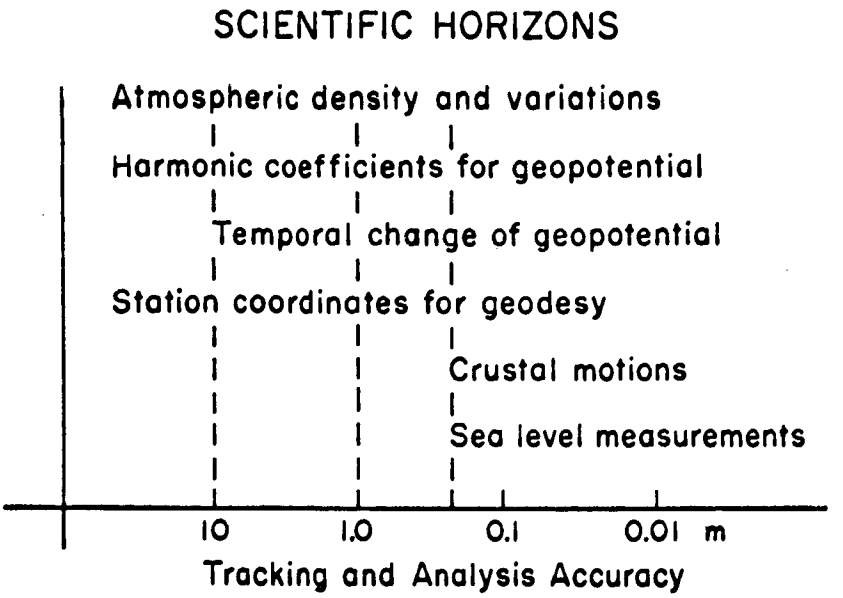


Figure 1. Scientific horizons.

Many topics must be examined as candidates for future scientific investigations using tracking data. Of course, the atmosphere remains an important subject, particularly in view of the relation to other magnetosphere phenomena. Also, a maximum of solar activity is approaching, and associated atmospheric changes must be studied.

Presumably, more accurate observations can lead to more accurate station coordinates. On the one hand, this has a direct bearing on the accuracy of mapping (Sherman, 1966). On the other hand, this may open the way for the detection of dynamical processes within our earth. Previous observational accuracies have been inadequate for the detection of any station displacements due to crustal motions or continental drifts (Markowitz, Stoyko, and Fedorov, 1964). This subject must be reexamined.

An important consequence of an accurate geopotential representation is the insight it provides into the internal constitution of the earth (Strange, 1966; Wang, 1966; Kaula, 1966). New determinations of the geopotential will be possible with new data. These will allow more penetrating interpretations and the study of such phenomena as earth tides (Newton, 1965; Kozai, 1965, 1966).

The interface of satellite geodesy with oceanography is a popular subject of discussion these days. This is partly due to the growing interest from NASA's expanding applications program (Badgley, 1965). A panel on currents and the shape of the geoid gave 50 cm as the threshold accuracy for useful measurements of actual sea level (Stewart, 1965). Radars to make such measurements from a satellite are being developed by industry. Presumably, comparable orbital positions will be required.

With respect to analytical techniques and celestial mechanics, there is an interplay between treatments of earth satellites, lunar orbiters, and other spacecraft and bodies in the solar system. In the interest of economy, theoretical developments and computer programs might be done with enough generality to cover several of these cases.

The transmission of light through the atmosphere, particularly short pulses from flash tubes or lasers, may provide opportunities for research. This possibility may be particularly pertinent when the atmosphere becomes the limiting factor influencing tracking accuracies.

There is hope in some circles that laser tracking of a satellite at several earth radii may provide a check of relativistic mechanics and fundamental ideas about gravity. A project to launch a satellite for this purpose has been suggested.

Finally, as satellite observations and analysis increase in accuracy, there are many questions to be answered about the adequacy of international time keeping, polar-motion measurement, and even the value of fundamental constants. Statistical problems pervade the whole field.

A few of the topics for present and future scientific investigation are shown as examples in Figure 1. All the scientific horizons mentioned above are explored in greater detail in subsequent papers.

REFERENCES

BADGLEY, P. C.

1965. Introductory briefing. In Oceanography from Space, ed. by G. C. Ewing, Woods Hole Oceanographic Institution, Woods Hole, Mass., pp. 1-17.

GAPOSCHKIN, E. M.

1966a. A new determination of tesseral harmonics of the geopotential and station coordinates from Baker-Nunn observations. Trans. Amer. Geophys. Union (abstract), vol. 47, p. 47.

1966b. A dynamical solution for the tesseral harmonics of the geopotential and station coordinates using Baker-Nunn data. In Proceedings VIIth International COSPAR Space Sci. Symp., North-Holland Publ. Co., Amsterdam, in press.

JACCHIA, L. G.

1963. Variations in the earth's upper atmosphere as revealed by satellite drag. Rev. Mod. Phys., vol. 35, pp. 973-991.

1965. Static diffusion models of the upper atmosphere with empirical temperature profiles. Smithsonian Contr. Astrophys., vol. 8, no. 9, pp. 215-257.

1966. Properties of the upper atmosphere determined from satellite orbits. Discussion Meeting on Orbital Analysis, Proc. Roy. Soc., London, October, in press.

JACCHIA, L. G., AND SLOWEY, J.

1966. The shape and location of the diurnal bulge in the upper atmosphere. Smithsonian Astrophys. Obs. Spec. Rep. No. 207, 22 pp.

KAULA, W. M.

1966. Tests and combination of satellite determinations of the gravity field with gravimetry. Journ. Geophys. Res., vol. 71, pp. 5303-5314.

KÖHNLEIN, W. J.

1966. Corrections to station coordinates and to nonzonal harmonics from Baker-Nunn observations. In Proceedings VIIth International COSPAR Space Sci. Symp., North-Holland Publ. Co., Amsterdam, in press.

KOZAI, Y.

1964. New determination of zonal harmonics coefficients of the earth's gravitational potential. Smithsonian Astrophys. Obs. Spec. Rep. No. 165, 38 pp.

1965. Long-range analysis of satellite observations. Presented at Symposium on Satellite Geodesy, Athens, April.

1966. Determination of Love's number. In Proceedings Discussion Meeting on Orbital Analysis, Proc. Roy. Soc., London, October, in press.

LUNDQUIST, C. A.

1966. Satellite geodesy at the Smithsonian Astrophysical Observatory. Presented at American Congress of Surveying and Mapping and the American Society of Photogrammetry Convention, Washington, D.C., Submitted to Surveying and Mapping.

LUNDQUIST, C. A., AND VEIS, G., EDS.

1966. Geodetic Parameters for a 1966 Smithsonian Standard Earth. Smithsonian Astrophys. Obs. Spec. Rep. No. 200, 683 pp.

MARKOWITZ, W., STOYKO, N., AND FEDOROV, E. P.

1964. Longitude and latitude. In Research in Geophysics, vol. 2, ed. by H. Odishaw, MIT Press, Cambridge, Mass., pp. 149-162.

NEWTON, R. R.

1965. An observation of the satellite perturbation produced by the solar tide. Journ. Geophys. Res., vol. 70, pp. 5983-5989.

SHERMAN, J. C., Chairman

1966. Mapping, imagery, and data processing. In Spacecraft in Geographic Research, Natl. Acad. Sci. Natl. Res. Council, Washington, D.C., pp. 94-101.

STEWART, R. W.

1965. Recommendations of the panel on currents. In Oceanography from Space, ed. by G. C. Ewing, Woods Hole Oceanographic Institution, Woods Hole, Mass., pp. 19-20.

STRANGE, W. E.

1966. A comparison of the satellite derived gravity field with surface gravity (abstract). In Proceedings VIIth International COSPAR Space Sci. Symp., North-Holland Publ. Co., Amsterdam, in press.

VEIS, G.

1966a. Results from geometric methods (abstract). In Proceedings VIIth International COSPAR Space Sci. Symp., North-Holland Publ. Co., Amsterdam, in press.

1966b. Geodetic interpretation of the results (abstract). In Proceedings VIIth International COSPAR Space Sci. Symp., North-Holland Publ. Co., Amsterdam, in press.

WANG, C.-Y.

1966. Earth's zonal deformations. Journ. Geophys. Res., vol. 71, pp. 1713-1720.

WHIPPLE, F. L.

1966. On the satellite geodesy program at the Smithsonian Astrophysical Observatory. In Proceedings VIIth International COSPAR Space Sci. Symp., North-Holland Publ. Co., Amsterdam, in press.

WHIPPLE, F. L., AND HYNEK, J. A.

1958. Optical and visual tracking of artificial satellites. In Proceedings VIIIth International Astronautical Congress, Springer-Verlag, Vienna, pp. 429-435.

ESTIMATION OF SCIENTIFIC PARAMETERS

Henry D. Friedman[1]

The reduction of large samples of data to a few meaningful numbers generally falls (implicitly or explicitly) into the province of statistical estimation theory. Sometimes the data consist of direct measurements (with error) of one or more parameters of interest; in this case, simple averages may be used to estimate the values of the parameters. Sometimes a parameter can be estimated only indirectly from the observed data, and a relationship between parameter and data must be assumed and used.

The basic measurements of the satellite-tracking process are of the second variety. The satellite is in motion, described by a mathematical model of the orbit. The parameters of the model are not "observables" in the usual sense; hence their values can be estimated only indirectly.

Perhaps the best known, and most useful, method of estimation is that of least squares. We assume parameters α, β, σ, ... and observables $y_i = y_i(\alpha, \beta, \sigma, \ldots)$. The functional relationship $y_i = y_i(\alpha, \beta, \sigma, \ldots)$ is a mathematical model of the physical problem. In observing y_i we make an observational error or measurement error ϵ_i, so that we measure instead $z_i = y_i + \epsilon_i$. The method of least squares consists of finding estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}$, ... of α, β, σ, ... such that $\Sigma (z_i - y_i)^2$ is minimized, where the sum extends over all values of i.

A more basic derivation of the method of least squares results from the assumption that the errors ϵ_i are independent, identically distributed, normal random variables. In this probabilistic framework, one seeks the most

---

[1]Chief, Data Processing Department, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts; now at General Learning Corporation, Washington, D. C.

probable values of the parameters. The maximum-likelihood method is employed, and one attempts to maximize the joint probability density function

$$L = \prod_{\text{all } i} \frac{1}{\sqrt{2\pi}\sigma} e^{-\epsilon_i^2/2\sigma^2} .$$

Since also

$$L = c\, e^{-\frac{1}{2\sigma^2}\Sigma\epsilon_i^2} ,$$

it is clear that maximizing L is equivalent to minimizing

$$\Sigma\epsilon_i^2 = \Sigma (z_i - y_i)^2 .$$

Thus, the maximum-likelihood method, applied to normal or Gaussian errors, reduces to the method of least squares. The method was originated by Gauss, who applied it to the determination of the most probable orbit from a set of observations.

The discussion above applies to the simplest case of independent, identically distributed errors. Frequently, it is known that the errors are dependent and have unequal variances or weights, etc. In such cases, it is necessary to derive afresh the appropriate least-squares solution to the parameter-estimation problem. This is conveniently accomplished through the maximum-likelihood formulation. We can write a probability expression for the joint error distribution $L = f(\epsilon_1, \epsilon_2, \ldots)$, which contains the known or unknown probability parameters (such as means, variances, and covariances). If $\epsilon_i = z_i - y_i$ is substituted into this expression, and if L is maximized through the appropriate choice of parameter estimates α, β, σ, ..., then these estimates form the desired least-squares result.

SATELLITE TRACKING WITH A LASER

Carlton G. Lehr[1]

A laser system becomes practical for tracking a satellite when the satellite is equipped with a retroreflector on its surface. The retroreflectors currently in use are arrays of several hundred quartz corner cubes. Such reflectors are on the satellites given in Table 1.

Table 1. Retroreflector-equipped satellites

Satellite	Perigee (Mm)	Apogee (Mm)	Inclination (deg)	Stabilization
64064-01	0.89	1.09	80	magnetic
65032-01	0.94	1.32	41	magnetic
65089-01	1.12	2.27	59	gravitational

The launch of another retroreflecting satellite, the French D1-C, is expected during 1967. This satellite will be magnetically stabilized, and its planned orbit has an inclination of 40°, a perigee of 0.6 Mm, and an apogee of 1.8 Mm (COSPAR, 1966). The detection of a laser return from a large passive satellite such as Echo 2 is marginally possible, but the practical use of laser systems will probably be confined to the satellites with retroreflectors.

[1]Staff Engineer, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts.

A laser can be used to illuminate a retroreflecting satellite when it is in the earth's shadow (Iliff, 1965; Anderson, Lehr, Maestre, Halsey, and Snyder, 1966). Then the satellite can be photographed against a star background by a Baker-Nunn or other camera just as if the object were in sunlight.

A laser system can be used in another manner, however, to track satellites, one that yields data whose geodetic significance is probably greater. In this mode of operation, a short pulse is transmitted to the satellite and reflected back to earth. The two-way transmission time of the pulse is measured by a time-interval counter that is incorporated in the system (Plotkin, 1964, 1965; Bivas and Morael-Courtois, 1966; Anderson, Lehr, and Maestre, 1965; Lehr, 1966; Lehr, Maestre, and Anderson, 1966). Such a measurement gives the range of the satellite normalized to the velocity of light. This normalization seems not to affect the geodetic usefulness of the data. The following three agencies have laser systems of this type that are currently in routine operation:

A. Goddard Space Flight Center, Greenbelt, Maryland.

B. Centre National de la Recherche Scientifique at l'Observatorie de Haute-Provence, France.

C. Smithsonian Astrophysical Observatory's station in Las Cruces, New Mexico.

Each of the three systems has a laser power of tens of megawatts and a pulse duration of tens of nanoseconds. The NASA and French systems operate at pulse repetition rates of 1 and 4 sec^{-1}, respectively, whereas the SAO system is 1 min^{-1}. Consequently, the first two systems make several hundred range measurements during a single pass of a satellite, while the SAO system makes only a few measurements.

The systems with the higher repetition rates have an advantage over the SAO system when a single station is operated. This advantage decreases when a network of stations, such as that planned by SAO, is established. In fact, for 12 geographically distributed laser stations, observations at rates greater than 1 min^{-1} would probably not increase the accuracy of the computed orbits.

The present SAO laser system uses predicted values of a satellite's azimuth and altitude to acquire the satellite in sighting telescopes attached to the surplus naval gun mount that carries the laser. Once the satellite is acquired, it is tracked visually by two observers, one following the satellite's azimuth and the other its elevation. At the same time, the laser receiver – a converted searchlight using a photomultiplier as a detector – also tracks the satellite, by means of visual operation of a third observer who uses a joy-stick velocity control. The laser is fired at known epochs by the crystal clock at the station. The elapsed time corresponding to the two-way travel of the laser pulse is obtained from an electronic time-interval counter. Electrical time delays in the system are subtracted from a range calibration made on a target at an accurately surveyed distance. [2]

A calculation of the expected error in the laser measurements presents many difficulties; consequently, it seems best at present to estimate it from experimental data that have already been obtained.

Preliminary determinations of the typical accuracy of a laser system have been made in the following ways. The several hundred data points for a single pass have been fitted to smooth curves at CNRS and GSFC. For each case, they find a standard deviation of 1. 5 m in the range data. This value is consistent with the accuracy of their range counter (±10 nsec) and the

[2] Editors' Note: The installation of a new laser system is planned for SAO's observatory at Mt. Hopkins, Arizona, during 1967. This system will be used to test the operating concepts that currently seem feasible and desirable when a network of geographically distributed laser stations is considered. The new system will have the laser and the receiver on the same mount, one that can be positioned by a single man to within 1/2 arcmin. The laser will be an oscillator-amplifier chain operating at 694. 3 nm and having a pulse length of 10 nsec and a power output of 500 Mw. The receiver will be a telescope of 20-inch aperture. It is expected that the mount can be repositioned and the laser fired every 30 sec. The beamwidths of both laser and telescope will be adjustable from 2 arcmin to 20 arcmin, a variation that should compensate well for errors in predictions of the satellite's azimuth and elevation.

duration of their laser pulse (20 to 30 nsec). The range data obtained by the SAO system have been included with precisely reduced Baker-Nunn observations in the computation of orbits through the Differential Orbit Improvement Program. The residuals of the laser observations reflect the accuracy of the orbits for the particular GEOS orbits that were computed. These residuals were of the order of tens of meters. They show a consistency between the Baker-Nunn data and the laser data, but they reflect the accuracy of the camera rather than that of the laser. They show that no large unexpected bias errors in the laser measurements are apparent. The standard deviation of a range measurement, obtained from an error analysis of the present SAO system, is 3 m. This value comes mainly from the laser's pulse duration of 60 nsec and the assumption that the counter would be equally likely to record a weak returned signal at any point within this 60-nsec interval.

The maximum satellite range measured with the SAO system is 3.9 Mm, a value that appears to be consistent with maximum ranges achieved with the other systems. An analysis of the signal strengths of many returns has shown that the signals are at least 5 to 10 db below calculated values, and many returns are significantly weaker than this. In fact, the returns for a given range show variations of about 25 db from one measurement to another. The cause of these variations has not yet been determined. Consequently, future laser systems must be designed for greater performance than that indicated by calculations based on the present use of the radar-range equation.

Although measurements from laser systems like the present ones should improve the orbits obtained from camera observations, measurements from improved laser systems will have important applications of still greater significance. It is possible to estimate the accuracy of a laser system that consists of the best components now available commercially. Of course, such an estimate is based only on known sources of error; experiments are necessary to ensure that the ultimate accuracy can be achieved in practice.

The duration of the laser pulse is a source of error, particularly when the returning signal is weak and only the presence or absence, but not the structure, of the returned pulse can be determined. The minimum pulse duration of lasers now commercially available is about 10 nsec. The pulse shape of such a laser is nearly Gaussian, with the half-power points 10 nsec apart. The corresponding standard deviation is 4 nsec, a value equivalent to an error in range measurement of $\sigma_1 = 0.6$ m.

If the returned signal is weak, the phototube generates only a few electrons during the pulse's duration. Consequently, the shape of the pulse is not reconstructed by the receiver, and a specific epoch within the pulse cannot be located with an error less than about σ_1. If, on the other hand, the returned signal is strong, the structure of the pulse can be determined. Then the transmitted and received pulses can be correlated and an increase in the accuracy of the range measurement can be obtained. In estimating the system's accuracy, correlation techniques will not be considered here, and the value $\sigma_1 = 0.6$ m is accepted as a minimum one for the laser pulse.

Another source of error is associated with the time-interval counter that measures the two-way travel time of the laser pulse. A counter, accurate to ±0.1 nsec, is offered commercially; the corresponding standard deviation is $\sigma_2 = 0.004$ m. The earth's atmosphere introduces an error into the range measurements, but with corrections based on local values of temperature and barometric pressure, the residual error is only $\sigma_3 = 0.15$ m. Through combination of the various errors, the total system error takes the following value:

$$\left(\sigma_1^2 + \sigma_2^2 + \sigma_3^2\right)^{1/2} = 0.6 \text{ m} \quad .$$

In addition to the precision of its range measurements, a laser system has another, as yet unexploited, advantage in satellite tracking — the increased distribution achieved by obtaining observations when the satellite is in the earth's shadow or when the sky is bright. The fact that the laser is an active system makes it independent of natural illumination of the satellite. The

spectral purity and the narrow divergence of the returning laser beam facilitate its detection in the presence of ambient light. This mode of operation, which is independent of visual acquisition of the satellite, requires predictions of the satellite's position to within the width of the laser's beam (i.e., to a few minutes of arc). Since field observations with the Baker-Nunn camera are accurate to about 1 arcmin, predictions based on such observations, if kept current, should be satisfactory.[3] In this connection, we note that lasers are available at the gigawatt-power level; consequently, the transmitted beam could be broadened somewhat to compensate for prediction errors and still provide sufficient power density for a detectable return.

[3] Editor's Note: A new prediction program, written for laser tracking, has been tested as follows: GEOS orbits based on 294 field-reduced observations for the 2-week period prior to March 1, 1966, were extrapolated to obtain satellite predictions for the 6-day period from March 5 to March 10, 1966. These predictions were then compared with the corresponding field observations. The comparison showed that the prediction error was about 2 arcmin near the beginning of the 6-day period and about 6 arcmin near the end of this period. The results indicate that predictions of sufficient accuracy can be obtained. These can be used to position the laser mount. Visual tracking should not be necessary.

BIBLIOGRAPHY

ANDERSON, P. H., LEHR, C. G., AND MAESTRE, L. A.

1965. Use of a laser for satellite range measurements. Smithsonian Astrophys. Obs. Spec. Rep. No. 190R, 26 pp.

ANDERSON, P. H., LEHR, C. G., MAESTRE, L. A., HALSEY, H. W., AND SNYDER, G. L.

1966. The two-way transmission of a ruby-laser beam between earth and a retroreflecting satellite. Proc. IEEE, vol. 54, pp. 426-427.

BIVAS, R., and MORAEL-COURTOIS, N.

1966. Détermination de l'orbite du satellite GEOS-A au moyen d'un télémètre à laser à partir d'une station. Compte Rend., vol. 262, pp. 935-937.

COSPAR

1966. COSPAR Information Bull. No. 32, Regular Issue, pp. 43-45.

ILIFF, R. L.

1965. Photographing satellite reflected laser pulses for geodetic stereo triangulation. Journ. Geophys. Res., vol. 70, pp. 3505-3508.

LEHR, C. G.

1966. Satellite tracking with a laser. Smithsonian Astrophys. Obs. Spec. Rep. No. 215, 30 pp.

LEHR, C. G., MAESTRE, L. A., and ANDERSON, P. H.

1966. Measurements of satellite range with a ruby laser. Smithsonian Astrophys. Obs. Spec. Rep. No. 211, 27 pp.

LENGYEL, B. A.

1966. Introduction to Laser Physics. John Wiley and Sons, Inc., New York, 311 pp.

PLOTKIN, H. H.

1964. The S-66 laser satellite tracking experiment. In Quantum Electronics, III, vol. 2., ed. by P. Grivet and N. Bloembergen, Columbia Univ. Press, New York, pp. 1319-1332.

1965. Tracking of the Beacon-Explorer Satellites with laser beams. COSPAR Information Bull. No. 29, Regular Issue, pp. 18-21.

ROSS, M.

1966. Laser Receivers. John Wiley and Sons, Inc., New York, 405 pp.

SMITH, W. V., and SOROKIN, P. P.

1966. The Laser. McGraw-Hill Book Co., New York, 448 pp.

THE EFFECT OF THE ATMOSPHERE ON LASER RANGE DETERMINATION

Henry G. Horak[1]

A laser pulse, or beam, passing through the earth's atmosphere is subject to various disturbances. The flux density diminishes and fluctuates as the beam penetrates; the direction varies; the speed of light varies; there is loss of coherence; and the polarization is altered. Insofar as range measurements are concerned, the effects of extinction, which include both absorption and scattering, and refraction are most important. Extinction affects the strength of the return signal, but it does not enter into the calculations since either sufficient flux gets through or it does not. If we assume a detectable signal, the accuracy of the range measurement is limited by instrumental effects and by the degree to which the refraction can be calculated.

The apparent range R_a of a satellite is obtained by measuring the length of time T required for a laser pulse to travel from the transmitter to the retroreflector and back to the receiver; thus,

$$R_a = c \int_0^{T/2} dt \quad , \tag{1}$$

where t is time and c the speed of light in a vacuum. The path of the laser beam in the atmosphere is a curve, and we shall measure arc length s along this curve from an origin O at the transmitter. The geometric path length from the origin to the satellite is designated

$$R_g = \int ds \quad . \tag{2}$$

[1] Astronomer, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts; regularly at the University of Kansas, Lawrence, Kansas.

We shall let R represent the straight-line range from transmitter to satellite. The error ΔR_a of the apparent range can be expressed as follows:

$$\Delta R_a = \int c\, dt - R = \int n(s)\, ds - R \quad ,$$

where n(s) is the index of refraction of air at the point s of the beam path, or

$$\Delta R_a = \int [n(s) - 1]\, ds + \left[\int ds - R \right] \quad . \tag{3}$$

The first integral denotes the refractivity error ΔR_n; the second, the geometric error ΔR_g. These errors can be computed by using the basic equation of geometric optics (Synge, 1937):

$$\frac{d(n\vec{T})}{ds} = \nabla n \quad , \tag{4}$$

where $\vec{T}$ is the unit tangent along the ray. We must know the variation of the index of refraction with height above the earth's surface before the solution of the equation can be attempted. Owing to the diligent work of many people, we are now provided with tables of model atmospheres, and the integration of the above equation can be carried out. Such calculations have been extensively made for short radio waves ($\lambda \sim 2$ cm). Table 1 (Bean and Thayer, 1963) gives typical and extreme values of range errors for targets beyond the atmosphere (heights above 70 km) for various observed angles of elevation ϵ_0. The small values of the geometric error are noteworthy, showing that the main correction is due to the varying speed of light along the ray path. We can obtain range accuracy to better than 1 m for $\epsilon_0 = 1°$, and 0.06 m for $\epsilon_0 = 90°$ by using an empirical formula into which is substituted the measured value of the surface index of refraction n_0 (Freeman, 1964). For visible light ($n_0 \sim 1.00027$) it should be possible to do even better than this.

The atmosphere exhibits turbulent motions, especially in the daytime, and these produce random fluctuations in the index of refraction, which in turn create, among other effects, range variations and deviations in the direction of the beam. It is reasonable to expect turbulence to be confined mainly to the surface layer, where diurnal heating and terrain irregularities

Table 1. Typical and extreme values of range errors

Elevation angle ϵ_0		Typical $n_0 = 1.000320$			Extreme $n_0 = 1.000400$		
		ΔR_n	ΔR_g	ΔR_a	ΔR_n	ΔR_g	ΔR_a
Milliradians	Degrees	Meters			Meters		
0	0	100.	10.	110.	165.	60.	225.
20	1	62.5	2.5	65.	73.	4.5	78.
50	3	38.1	0.7	39.	43.	1.	44.
100	6	22.3	0.14	22.	24.8	0.2	25.
200	11	11.9	0.02	12.	13.	0.3	13.
500	30	5.0	0.001	5.	5.5	0.002	5.5

$\lambda = 1.85$ cm; n_0 = surface value of the index of refraction.

result in ascending and descending air currents, and to those higher layers characterized by temperature inversions. However, there is a lack of observational data, for it is difficult to measure turbulence directly; of interest in this connection is the use of radar to track a rising superpressure balloon whose motion becomes erratic upon encountering turbulence (Endlich, Mancuso, and Davies, 1966). In any event, it is highly desirable to correlate turbulent conditions with the synoptic weather situation.

The basic theoretical study of the effect of a turbulent medium on the passage of a light ray is due to Chandrasekhar (1952). The following assumptions are made: 1) variations in the index of refraction are small, so that the angular deviations of a ray are small; 2) the ray path is calculated by the equation of geometric optics (equation (4)); and 3) the turbulent motions are isotropic. Chandrasekhar then shows that

$$\langle \Delta^2 R \rangle^{1/2} = (4.21 \times 10^2)(s r_0)^{1/2} \left(\overline{\Delta^2 n}\right)^{1/2} \text{cm} \quad , \tag{5}$$

and

$$\langle \Delta^2 \theta \rangle^{1/2} = (1.74 \times 10^7)\left(\frac{s}{r_0}\right)^{1/2} \left(\overline{\Delta^2 n}\right)^{1/2} \text{arcsec} \quad , \tag{6}$$

where $\langle \Delta^2 R \rangle^{1/2}$ is the root-mean-square fluctuation in the range, $\langle \Delta^2 \theta \rangle^{1/2}$ is the root-mean-square fluctuation in the deviation angle θ (the angle the emergent ray makes with the original direction), s is the length of the turbulent medium expressed in units of 10^4 cm, r_0 is the microscale (the "average" size of an eddy) expressed in units of 10 cm, and $\left(\overline{\Delta^2 n}\right)^{1/2}$ is the root-mean-square fluctuation in refractive index. Now an increment of 1° K in temperature corresponds to an increment of about 10^{-6} in the index of refraction of air, so that for reasonable values of s and r_0 the size of the root-mean-square fluctuation in range is small compared to 0.5 m, the order of magnitude of the uniform refraction error. If we assume the extreme conditions s = 100 km, r_0 = 10 m, and $\left(\overline{\Delta^2 n}\right)^{1/2} = 10^{-5}$, then $\langle \Delta^2 R \rangle^{1/2} \sim 1$ cm; however, $\langle \Delta^2 \theta \rangle^{1/2} \sim 2°$. We can therefore hardly expect turbulence to be of much importance insofar as range measurements are concerned. Of course, if the eddies are far from the receiver, the beam may occasionally deviate enough

to miss the receiver; this suggests (along with energy considerations) a large aperture for the receiver. It is also clear that turbulence is much more serious in communication problems, where a rather high degree of coherence must be maintained in the laser beam. Further discussion can be found in Chandrasekhar (1952, 1953), Robertson (1940), Tatarski (1961), Hufnagel and Stanley (1964), Beckmann (1965), Hodara (1966), Fried and Meyers (1965), Reiger (1962), and Meyer-Arendt and Emmanuel (1965).

For range measurements it is advisable to use a laser wavelength that is not close to a prominent atmospheric absorption feature. An atlas of these features from λ 6326 Å to λ 2.44 μ has been compiled by Long (1963). He also points out that the ruby-laser wavelength varies with temperature from λ 6934 Å at 88° K to λ 6943 Å at 300° K, so that the ruby laser can be temperature tuned to avoid atmospheric absorption lines. (See also Curcio, Drummeter, and Knestrick, 1964; Elterman, 1964; Plass, 1966.)

Laser light has more penetrating capability than ordinary noncoherent light. Carrier and Nugent (1965) investigated this by measuring the volume scattering functions and extinction coefficients of natural fogs for both coherent and noncoherent light sources (λ = 6330 Å). Both polarized and nonpolarized sources were used. Theoretical studies suggest the use of intensity-dependent terms in the extinction coefficient at least for high intensities (Condell, 1964; Condon, 1964).

Daylight measures of range can be seriously impaired by the atmospheric scattering of sunlight. The receiving sensor cannot distinguish between laser light reflected from the satellite and sunlight of the same wavelength scattered into the line of sight. However, the orbit of the satellite is assumed to be known sufficiently well for aiming purposes, so that it might be possible to estimate the time of arrival of the return signal, and, even under rather marginal conditions, to distinguish it from the sunlight "noise." The feasibility of making daylight measures is discussed by Lehr (1966), who concludes that with available equipment such observations lie within the realm of attainment.

ANNOTATED BIBLIOGRAPHY

BEAN, B. R., AND THAYER, G. D.

1963. Comparison of observed atmospheric radio refraction effects with values predicted through the use of surface weather observations. Radio Sci. Journ. Res. NBS, vol. 67D, pp. 273-285.

BECKMANN, P.

1965. Signal degeneration in laser beams propagated through a turbulent atmosphere. Radio Sci. Journ. NBS, vol. 69D, p. 629.

This excellent paper is a thorough, but concise, discussion using both geometric and wave optics.

CARRIER, L. W., AND NUGENT, L. J.

1965. Comparison of some recent experimental results of coherent and incoherent light scattering with theory. Appl. Opt., vol. 4, pp. 1457-1462.

CHANDRASEKHAR, S.

1952. A statistical basis for the theory of stellar scintillation. Monthly Notices Roy. Astron. Soc. London, vol. 112, pp. 475-483.

This is the basic paper on the subject, but it requires some preliminary study of the theory of isotropic turbulence.

1953. Some aspects of the statistical theory of turbulence. Proc. 4th Symp. Appl. Math., Amer. Math. Soc., p. 1.

A discussion of the theory of turbulence is presented, which will serve as background material for Chandrasekhar, 1952.

CONDELL, W. J.

1964. High-intensity propagation through absorptive or amplifying media. Journ. Opt. Soc. Amer., vol. 54, pp. 1166-1167.

CONDON, E. U.

1964. Intensity-dependent absorption of light. Proc. Natl. Acad. Sci., vol. 52, pp. 635-637.

CURCIO, J. A., DRUMMETER, L. F., AND KNESTRICK, G. L.

1964. An atlas of the absorption spectrum of the lower atmosphere from 5400 Å to 8520 Å. Appl. Opt., vol. 3, pp. 1401-1409.

Microphotometer traces are shown of the absorption spectrum of an uncontaminated 16-km path in the lower atmosphere; the resolution is ~0.2 Å.

ELTERMAN, L.

1964. Parameters for attenuation in the atmospheric windows for fifteen wavelengths. Appl. Opt., vol. 3, pp. 745-749.

ENDLICH, R. M., MANCUSO, R. L., AND DAVIES, J. W.

1966. Techniques for determining a world-wide climatology of turbulence through use of meteorological data. Stanford Res. Inst. Sci. Rep. 1, iv, 65 pp.

FREEMAN, J. J.

1964. Real time compensation for tropospheric radio refractive effects on range measurements. NASA Rep. CR-109, 18 pp.

FRIED, D. L., AND MEYERS, G. E.

1965. Atmospheric optical effects — polarization fluctuation. Journ. Opt. Soc. Amer., vol. 55, pp. 740-741.

HODARA, H.

1966. Laser wave propagation through the atmosphere. Proc. IEEE, vol. 54, pp. 368-375.

A rather elementary, but effective, pile-of-plates model of turbulence is used to derive expressions for beam scanning, phase variation, beam cross-section change, amplitude and frequency modulation, and polarization fluctuation.

HUFNAGEL, R. E., AND STANLEY, N. R.

1964. Modulation transfer function associated with image transmission through turbulent media. Journ. Opt. Soc. Amer., vol. 54, pp. 52-61.

This paper is concerned with those image-degrading effects of turbulence that remain after the image has been averaged over a long time.

LEHR, C. G.

1966. Satellite tracking with a laser. Smithsonian Astrophys. Obs., Spec. Rep. No. 215, 30 pp.

LONG, R. K.

1963. Absorption of laser radiation in the atmosphere. Ohio State Univ. Res. Foundation, AD410571.

Also refer to "Atmospheric attenuation of ruby lasers," Proc. IEEE, vol. 51, pp. 859-860.

MEYER-ARENDT, J. R., AND EMMANUEL, C. B.

1965. Optical scintillation; a survey of the literature. Tech. Note 225, NBS, 140 pp.

PLASS, G. N.

1966. The absorption of laser radiation along atmospheric slant paths. Appl. Opt., vol. 5, pp. 149-154.

This is a theoretical study of the effects of atmospheric absorption on laser radiation.

REIGER, S.

1962. Atmospheric turbulence and the scintillation of starlight. Rand Corp. Rep. No. R-406-R, 44 pp.

ROBERTSON, H. P.

1940. The invariant theory of isotropic turbulence. Proc. Camb. Phil. Soc., vol. 36, pp. 209-223.

The mathematics of isotropic tensors is applied to the theory of turbulence, and is useful background material for Chandrasekhar, 1952, 1953.

SYNGE, J. L.

1937. Geometrical optics: an introduction to Hamilton's method. Cambridge Tracts in Mathematics and Mathematical Physics, no. 37, Cambridge Univ. Press, Cambridge, England, 110 pp., Chap. 5.

A thorough introduction to Hamilton's method is given; the basic equation of optics is derived from Fermat's principle.

TATARSKI, V.

1961. Wave propagation in a turbulent medium. Russian trans. by R. A. Silverman. McGraw-Hill Book Co., New York, 285 pp.

PRESENT AND FUTURE RESEARCH ON THE UPPER ATMOSPHERE AT THE SMITHSONIAN ASTROPHYSICAL OBSERVATORY

Luigi G. Jacchia[1]

Our research on the upper atmosphere proceeds at four levels.

First, we are trying to contribute to the basic observational material by determining air densities from the drag analysis of suitable artificial satellites. Although a few instrumented satellites for the study of the neutral atmosphere have been launched, most of our information on upper atmosphere structure and variations has come so far from orbital drag analysis.

Second, we are trying to assemble all the available observational material and to analyze it in order to obtain an over-all phenomenological description of atmospheric properties.

Third, we are using the results of the analysis in the construction of semiempirical, descriptive models that will predict the density, and possibly other atmospheric parameters, as a function of height, geographic location, time of the day, and solar and geomagnetic indexes and other parameters.

Fourth, we are trying to construct a theoretical model of the atmosphere that, if the theory is any good, will, we hope, approach the features of the semiempirical model. In other words, we are trying to explain the observed features of the upper atmosphere on a physical basis, thereby gaining insight into the physical processes that are involved.

---

[1] Physicist, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts.

Although the main variations of the upper atmosphere were discovered and systematized in 1959 and 1960, new important details are coming to light every year.

In the variation with solar activity, we have found that the relation between atmospheric temperature and the decimetric solar flux is different according to whether it is determined over an 11-year sunspot cycle or over a 27-day solar rotation.

From our study of the variation with geomagnetic activity, we have found that the relation between the disturbance of the geomagnetic field and the temperature during magnetic storms is different from that observed during geomagnetically quieter periods. We also find that the heating of the atmosphere follows the geomagnetic disturbance with a lag of about 7 hours. This lag is somewhat smaller, 6 hours, in the auroral zones, where the heating is also intensified on occasions.

In the distribution of temperatures with respect to solar time and latitude, we have recently found significant anomalies at high latitudes. These are being actively investigated, since they promise to provide the clue to the heating mechanism of the upper atmosphere.

The present rise of solar activity is providing us with an opportunity to compare the behavior of ultraviolet heating and of corpuscular heating at different temperature levels; we hope this will lead to a better separation of the two effects. This type of study was not very successful during the decline in solar activity from 1959 to 1964, because satellite data in the early years (1959 to 1961) were less complete and less accurate.

Little is known about what causes the semiannual variation in the upper atmosphere. Electron densities in the F2 layer show a similar variation, which is strongly dependent on latitude. So far, satellite-drag data have not been sufficient to reveal any latitude dependence in the semiannual variation,

but it is hoped that future data will provide this opportunity. It has recently been found that the semiannual variation is very large at heights of the order of 1000 km, much larger than predicted by present-day models on the basis of thermospheric heating. This peculiarity deserves closer attention in future research.

Theoretical models, so far, have been based on diffusion equilibrium and have ignored lateral convection. Recent research seems to indicate that the diurnal variation must be accompanied by strong winds driven by pressure gradients, and that these winds may profoundly affect the geographic distribution of temperature. Also, the interaction between neutral gas and ions is found to affect the winds. The construction of a comprehensive model of the diurnal variation on a global scale presents formidable difficulties. We have been pursuing such a goal for nearly a year, but much remains to be done before we can be satisfied with our results.

CONTINENTAL DRIFT[1]

Ursula B. Marvin[2]

Earth history is subject to two radically different interpretations. The critical issue is whether or not the continents, with their long record of tectonic activity, have ever moved as massive units over the surface of the earth.

Many earth scientists believe that throughout geological time the continents have occupied fixed positions of latitude and longitude. They view the Precambrian shields as ancient continental nuclei, bordered by belts of successively younger rocks that through a series of geosynclinal-orogenic cycles have slowly built the continents outward to their present margins.

The alternative view holds that one or more continental masses have migrated over many degrees of latitude and longitude since the early Precambrian and that the present continents acquired their identity in a major breakup of these protocontinents about 200 million years ago. The subsequent drifting apart of the continents may have ceased in the mid-Tertiary or may be continuing today.

---

[1] Editors' Note: It is an interesting coincidence that during the period of the summer seminar a communique was received from the SAO station in Ethiopia, calling attention to the recent development of fissures on the flat floor of the main Ethiopian rift. P. A. Mohr and G. Gouin reported that about 90 km east-southeast of Addis Ababa the absolute amount of surface separation varies along the fissure line from a few centimeters to over 1/3 m. Similar fissures are reported to have developed during 1956. Interest in the topic of this seminar was stimulated by the reports from Ethiopia.

[2] Geologist, Smithsonian Astrophysical Observatory and Harvard University, Cambridge, Massachusetts.

Powerful arguments from geology and geophysics are marshalled in support of both interpretations. This paper reviews these arguments and, as no compromise is possible, concludes that the weight of the evidence available today favors continental drift.

The ultimate solution to the continental-drift problem may come from satellite geodesy. The SAO astrophysical observing stations are now located to an accuracy of 10 m, and the development of laser techniques will soon reduce the error to less than 1 m. This degree of accuracy will make possible the measurement of continental drift if it is occurring at the estimated rates of 1 to 10 cm year^{-1} (see page 67). Such precise measurement would require a network of laser stations that would make repeated determinations over a period of several decades, using great care to detect and correct for local crustal movements. In the end, a negative answer would prove little; present continental stability does not disprove past drift. If, however, the results were to show a definitive change in mean latitude or longitude of any landmass as a whole, these measurements would resolve one of the most spirited controversies in earth science.

1. THE CONTINENT PROBLEM

Continents exist in the crust of the earth, but theoretical calculations would not predict them. A spherical planet with a large gravitational field and a high internal temperature should differentiate into a heavy core surrounded by shells of decreasing density. The crust should, therefore, be uniform in thickness and composition, and, given the chemistry of the earth, it should be enveloped by a layer of water.

The degree of departure of the earth from this theoretical model is only beginning to be appreciated. Far from smooth and homogeneous, the crust is dominated by continents of irregular shape and distribution surrounded by ocean basins. These two topographic features are strikingly different in chemical composition, petrology, density, and age.

1.1 Topography and Isostatic Balance

The average height of the continents above sea level is 840 m; the average depth of the oceans is - 3800 m. Maximum crustal relief is 20 km, from Mt. Everest, 8848 m, to the Marianas Trench, -11,035 m. Gravity measurements had shown, by the turn of this century, that the topographic relief of the crust is largely in isostatic balance; i.e., mass in grams per square centimeter is equal under continents, mountains, and oceans, as though light materials are buoyed up in a denser substratum.

Seismic evidence confirmed this general picture when, in 1909, A. Mohorovičić discovered a major discontinuity where rocks of density 2.7 to 3.0 give way to heavier ones, density 3.3, in a pattern that reflects at depth the topographic relief of the earth's surface. This discontinuity is generally taken as the boundary between the crust and the mantle, and, with many exceptions, it occurs at an average depth of 5 km under the ocean floors, 35 km under the continents, and 60 km beneath high mountain ranges.

1.2 Chemical Composition and Density

The continents have the bulk composition of granodiorite, a light rock containing about 70% SiO_2 with an average density of 2.7. Mantle materials, below the Mohorovičić discontinuity, have a density of 3.3 to 3.4, which suggests that they are magnesium-rich silicate rocks such as perioditites or eclogites with 40 to 45% SiO_2. Recognizing the difference in composition between the crust and the mantle, Suess (1914) invented the names sal* (silica alumina) for the former and sima (silica magnesia) for the latter. This simple terminology seemed justified as long as the ocean basins were thought to be downfaulted blocks of continental rock. As early as 1910, however, gravity measurements threw doubt on this concept, and in the past two decades, extensive gravity and seismic data collected at sea have shown that the floors of the Atlantic and Indian Oceans, as well as that of the Pacific, are essentially basaltic (SiO_2 50%, density 2.95) in composition.

The sial proper is therefore strictly limited to the continents. The term sima is now frequently used for the oceanic basalts, and ultrasima is applied to the mantle. The restricted occurrence of sial suggests that it may never have formed a complete crustal shell but was always patchy in distribution. Recent seismic data indicate that the same is probably true of the oceanic basalts, which do not continue as a well-defined layer under the continents (Wasserburg, 1966).

1.3 Age

The continents and ocean floors differ in age as well as in density and composition. The continents contain rocks of all ages, from the most recent to very ancient Precambrian sediments that were metamorphosed 3600 million years ago. Continents were, therefore, in existence and undergoing erosion by about 4000 million years ago, only a few hundred million years after the

*The name sial was substituted for sal by Wegener (1924).

origin of the solar system (Donn, Donn, and Valentine, 1965). Since the early Precambrian, then, the continents have been persistent crustal units, partially downwarped and flooded at times but on balance, occurring as positive land masses with a continuous and well-preserved record of erosion, sedimentation, and diastrophism. The persistence of the continents can be explained only on the assumption that they have undergone continuous uplift at a rate that more than balanced the rate of erosion, which, operating alone, would have worn them down to sea level once and for all in about 25 million years (Kennedy, 1959).

In comparison with the continents, the ocean floors are geologically young. The oldest sediments in dredge or core samples from any ocean are late Jurassic, about 140 million years old (Wilson, 1965a). Basalt samples from oceanic ridges and seamounts range in age from about 1 to 60 million years. The oldest igneous rock from oceanic islands is 650 million years (Wasserburg, 1966), but this age may reflect an admixture of continental-type andesite and oceanic basalt.

The most striking evidence for the youth of the ocean floors, however, is that the total thickness of sediment is everywhere suprisingly low. The red clay and globigerina ooze typical of deep-ocean sedimentation is about 400 m thick in the Pacific and 500 m thick in the Atlantic. The deposition rates of the clay and ooze average about 1 mm and 10 mm per 1000 years, respectively. At a mean rate of 5 mm per thousand years the Pacific and Atlantic sediments could have accumulated in 80 to 100 million years, beginning in the late Cretaceous. Beneath the unconsolidated sediments lies a layer 1 to 2 km thick of higher density that is at least partially volcanic but may also contain lithified sediments (Gaskell, 1963). If this layer were wholly sedimentary it would indicate a Silurian age (about 400 million years) for the earliest ocean sediments. These figures are not precise, but they do suggest that the floors of the Atlantic and the Pacific are roughly the same age and that both are significantly younger than the continents.

Thus, a review of the evidence shows that the earth has no "crust" in the sense of a rock shell with characteristic properties. The outermost portion of the planet consists of two provinces, the continents and the ocean floors, which are fundamentally different in lithology and geologic history.

1.4 The Mantle

As the "crust" is inhomogeneous, so too is the mantle. Seismic, gravity, and heat-flow data, collected over the past decade, indicate that the mantle directly under the continents differs in density and composition from that under the ocean floors (MacDonald, 1963). Unlike basalts, granodiorites are rich in minerals that accommodate the large K, U, and Th ions, with the result that continental rocks are six times more radioactive than oceanic bedrock. Nevertheless, contrary to all predictions, the rate of heat flow from the ocean floors is, on the average, equal to that from continents and mountain ranges. If, as many geophysicists believe, this heat is generated by radioactivity, the equality of heat flow demands that the highly radioactive continental blocks be underlain by columns of mantle 300 to 500 km deep that are severely depleted in radioactive elements relative to the adjacent oceanic mantle. This suggests that the continents were formed by differentiation from the mantle rock directly beneath them (MacDonald, 1963).

Seismic measurements confirm density differences between continental and oceanic mantles and also several discontinuities, caused by phase or chemical layering, within the upper mantle (Knopoff, 1964; Anderson, 1965). The upper mantle therefore is not much more homogeneous a rock shell than the crust.

These gross inhomogeneities of the crust and upper mantle would seem to favor a stable configuration of continents and ocean basins throughout geological time. They clearly suggest that drift has not taken place along shallow levels such as the ocean floors or the Mohorovičić discontinuity so as to allow continental crust to slip over segments of oceanic mantle. Any

postulated theory of continental drift should account for or refute the evidence that the continents and ocean floors are each coupled with and genetically related to a separate column of mantle that retains its identity to depths of several hundred kilometers.

2. THE CONTINENTAL-DRIFT HYPOTHESIS

Earth history has been interpreted with a high degree of success in terms of stationary continents growing larger by the accretion of successively younger materials around the shield areas. All continents are believed to show this accretion to some extent, but the evidence, based on radiometric dating, is most complete for North America (Engel, 1963) (see Figure 1). The slow, marginal growth of continents, the inhomogeneous structure of the crust and upper mantle, and the worldwide development of the stratigraphic record all seem to favor, in the view of many earth scientists, a stable configuration of continents throughout geological time. Nevertheless, the hypothesis of continental drift is currently gaining wide acceptance.

Continental drift is an old concept that first acquired a large scientific following when it was proposed in 1912 by Alfred Wegener and expanded upon by Alexander Du Toit through the 1930's. Wegener had a broad command of natural sciences including geology, astronomy, and meteorology, and many of his pioneering concepts have recently proved to be correct. He was, for example, the first scientist to deduce that the ocean floors were not sunken continental blocks but that the continents and ocean floors were of different composition and density. He was also the first to outline in detail the apparent continuity of numerous structural and stratigraphic units in continents and islands that are separated by ocean basins. He postulated that, at the end of the Mesozoic, blocks of light, continental sial had broken from a former protocontinent and drifted apart through the heavier, oceanic sima. His theory attracted many geologists, particularly those who were puzzling over the distribution of sedimentary deposits and organisms in the Southern Hemisphere, but it was opposed by others, including those who clung to the belief that all oceans except the Pacific had sialic floors, an assumption that was finally disproved after World War II.

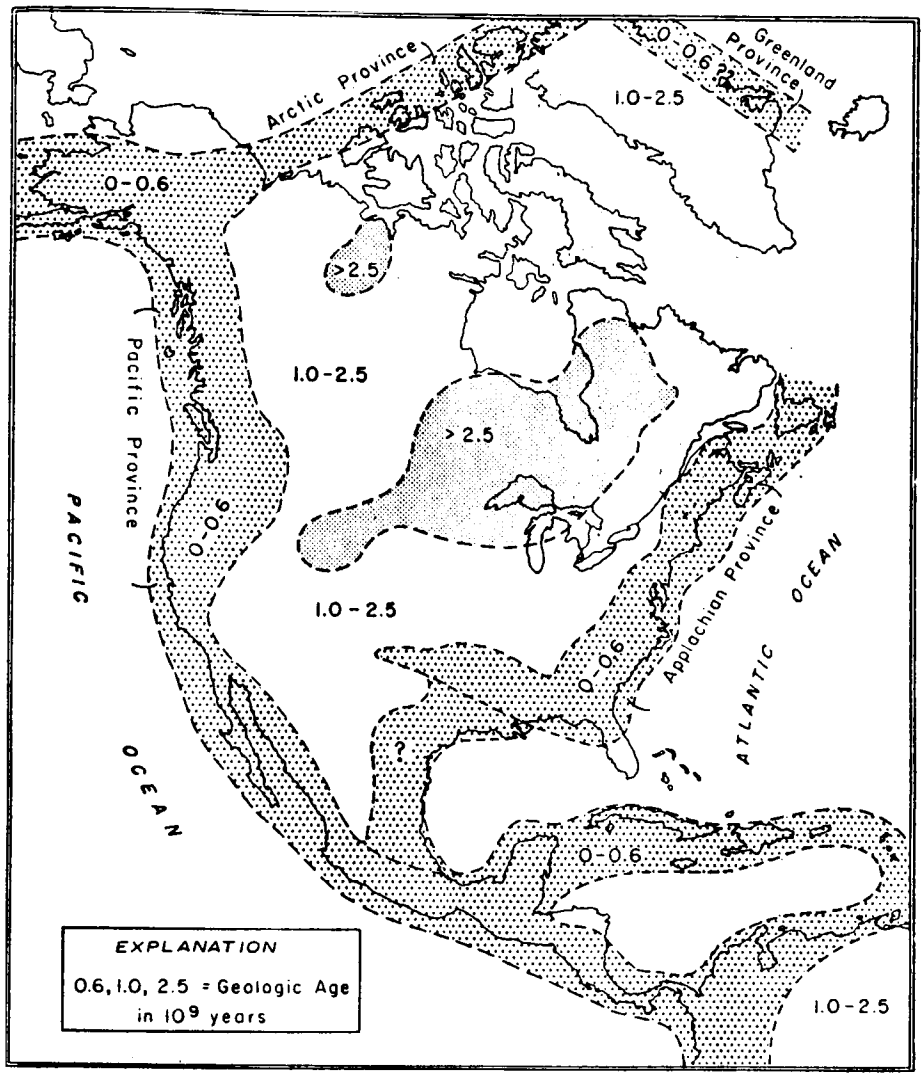


Figure 1. Continental accretion (from Engel, 1963, p. 145).

In 1928 the drift hypothesis suffered a loss of credibility and ceased to be an acceptable subject of geological investigation in many universities as a result of a symposium held under the chairmanship of W. A. J. M. van Waterschoot van der Gracht (1928). A rereading of the papers of the symposium makes it difficult to understand this overwhelmingly negative reaction, particularly as the chairman himself favored drift. The reaction was, however, preconditioned in large part by the opposition to continental drift expressed earlier by Jeffreys (1924), whose enormous prestige as a geophysicist proved decisive with many geologists. Jeffreys argued that the force of gravity is greater than any known tangential force acting within the earth, that the continental and oceanic crusts are strong enough to maintain topographic features such as Mt. Everest and the deep ocean trenches without slowly spreading out under the pull of gravity, and that they are therefore too strong to permit the horizontal drifting of sialic blocks through the sima. He also observed that the disequilibrium figure of the earth demonstrates the presence of a strong mantle. Jeffreys expanded upon his arguments through five editions of his book The Earth, and he is still, today, one of the strongest opponents of continental drift.

Over the past decade, however, the theory of continental drift has gained new vigor and new adherents as a result of advances in the studies of paleomagnetism and oceanography. Although many different configurations of ancient landmasses have been proposed, they all require a Mesozoic date for the breaking up and drifting apart of the present continents. Wegener (1924) believed the breakup was initiated about 70 million years ago, but current estimates range from about 250 to 150 million years ago.

Continental drift is postulated to explain the following:

A. The apparent fit of continents and of truncated structural trends separated by oceans — particularly the Atlantic and Indian Oceans.

B. The bizarre pattern of climatic zones in the late Paleozoic, as shown by the distribution of glacial deposits and fossil plants and animals.

C. The apparent migration of the magnetic poles along paths that differ for different continents, as shown by remnant magnetism in rocks older than mid-Tertiary.

D. The youth and topography of the ocean floors, and the linear patterns of magnetic anomalies paralleling the oceanic ridges.

2.1 The Continental Margins

The Atlantic shorelines of Africa and South America look like complementary pieces of a fractured block, and when they are matched along the 500-fathom line of the continental slopes, their fit is even better. North America and Europe are not so obviously paired, but a good fit can be made in the northern latitudes from Baffin Island to Greenland and Scandinavia. The northern and southern continents can then be joined so that the northwestern coast of Africa lies adjacent to North America from Florida to the Grand Banks (see Figure 2). This configuration requires the rotation of Spain into the Bay of Biscay, and the omission of Iceland, most of Central America, and the Caribbean Islands.

The eastern continents and Antarctica have been fitted by various schemes. Wegener assembled all the continents into a single large mass, which he called Pangaea (Figure 3). Du Toit believed there had been two primeval continents, Laurasia in the Northern Hemisphere, and Gondwanaland in the Southern. Others have postulated at least four early landmasses.

Although the pattern shown in Figure 2 requires some distortions and omissions, it is nevertheless so remarkably good a fit that some investigators see it as a compelling argument that these continents were formerly part of a massive block. Others maintain that continents could not possibly fit so well if they had been through a catastrophic event powerful enough to split open the crust and upper mantle and set seven continents drifting to different parts of the planet. Judging from the faults and folds of every geosyncline, they would predict major deformation and igneous activity along the newly rifted continental borders and the slabbing off of shoals of small sialic islands

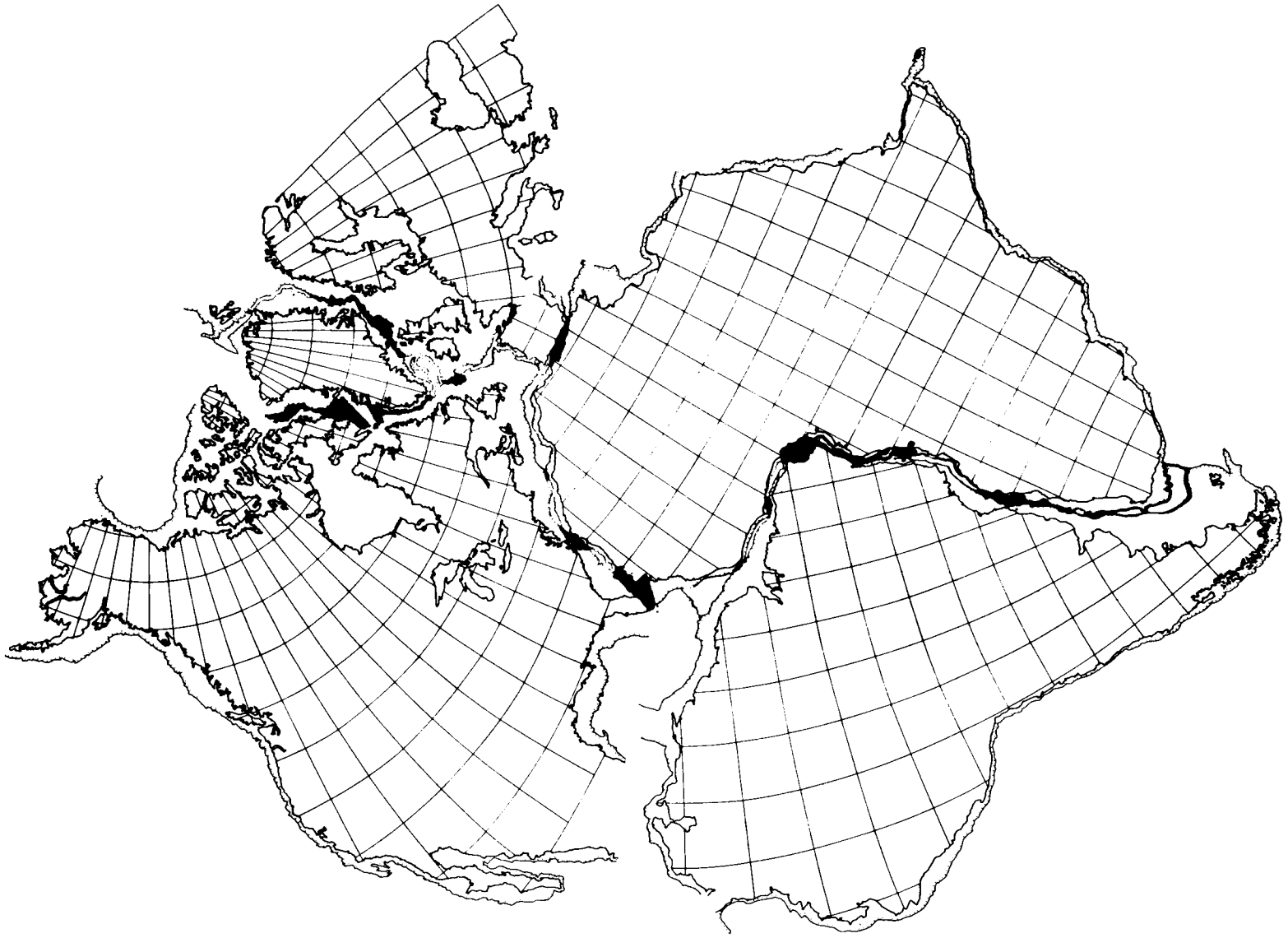


Figure 2. Fit of the continents around the Atlantic at the 500-fathom contour, transverse Mercator projection. Black areas are the overlap; the white areas between the dark lines represent the gaps (from Bullard, Everett, and Smith, 1965, p. 48).

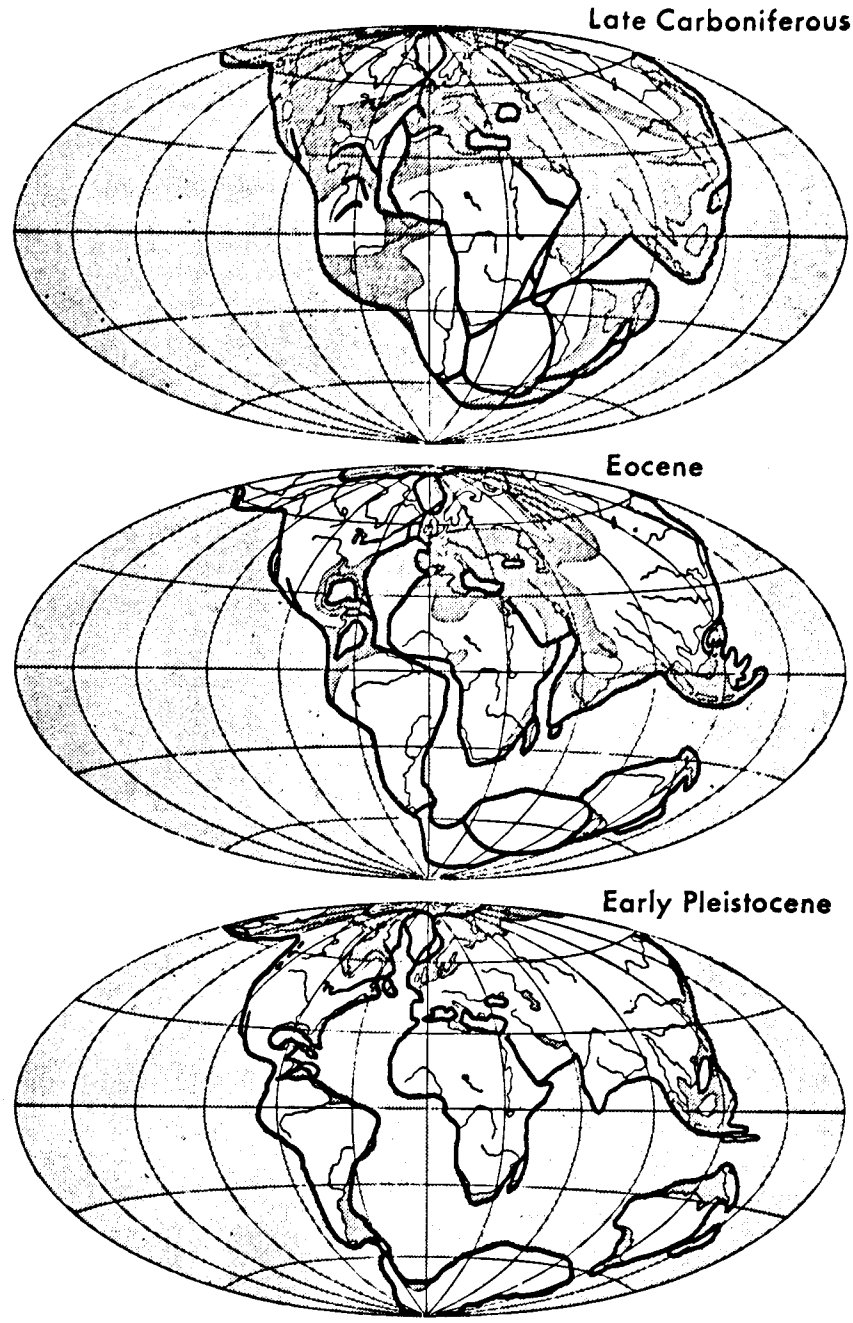


Figure 3. Wegener's reconstruction of the distribution of the continents during the periods indicated. Africa is placed in its present-day position to serve as a standard of reference. The more heavily shaded areas (mainly on the continents) represent shallow seas (from A. Wegener, 1924, p. 6).

(like the Seychelles and Madagascar). A close fit is thus regarded as either favorable or unfavorable to continental drift depending upon one's point of view. The fit of continental margins, however, unconfirmed by other evidence, is accepted as a critical criterion by no one.

Orogenic belts and lithologic contacts are abruptly truncated across their strike on both sides of the Atlantic and Indian Oceans. The Appalachian folded belt, for example, strikes out to sea in Newfoundland and has an apparent continuation in the Caledonian belt of Ireland and Scotland. Similarly, Paleozoic and early Mesozoic folding in the Cape geosyncline of South Africa is apparently matched in the Santa de la Ventana of Argentina. Prior to the development of radiometric dating methods, the equivalent ages and implied continuity of these and other units were often in question.

A current project of dating the rocks of Brazil and West Africa by both the K/Ar and the Rb/Sr methods is being carried on jointly by the geochronology laboratories at MIT, the University of São Paulo, and several other institutions. The results are beginning to demonstrate a matching of rock provinces and contacts across the Atlantic to a degree well beyond anything likely to occur by coincidence. Using Bullard, Everett, and Smith's (1965) reconstruction of the two continents (Figure 2), a mapped contact, striking toward the Atlantic in Ghana, between rock provinces 2000 million and 500 million years old was extended westward to a predicted location near São Luis in northeastern Brazil. Subsequent sampling and dating of the basement rocks around São Luis have established the presence of this contact where it was predicted (Hurley, Melcher, Rand, Fairbain, and Pinson, 1966).

Farther south along the Brazilian coast a second belt of 2000-million-year-old rocks matches another one in West Africa. Each of these series of rocks is similar to its counterpart across the Atlantic in lithology and ore deposits, as well as in age. The elegant fit of these contacts, in addition to that of the continental margins, makes a persuasive case for the former continuity of Africa and South America in the Precambrian and early Paleozoic.

That this continuity persisted through the Triassic is suggested by the apparent matching of the late Paleozoic sedimentary sections and the structural trends of Triassic dike systems in each continent. Furthermore, the oldest marine embayments, demonstrating the presence of an ocean, along the Atlantic coastlines of either Africa or South America are Jurassic, less than 180 million years old.

2.2 Permocarboniferous Climates

The distribution of glacial tillites of Permocarboniferous age is so irregular as to force a choice between:

A. Stationary continents with major climatic fluctuations independent of latitude, or

B. Climatic zoning by latitude with migrating continents.

Permocarboniferous tillites have been mapped in Brazil, Uruguay, Argentina, the Falkland Islands, Africa from the equator to the Cape of Good Hope, Madagascar, India, Australia, and Antarctica (see Figure 4a). Unlike the Pleistocene glaciers, which were confined to high latitudes and mountain ranges, Permian glaciers apparently occurred in the tropics at sea level. They are, in part, interbedded with marine deposits and coal seams. Except for India and a mountain range in Alaska, the Northern Hemisphere was evidently glacier-free, and warm and humid enough for the large-scale development of coal swamps in the same period. (The Permian tillite at Squantum, Massachusetts, was reclassified as a turbidite by Dott (1961).)

With the retreat of the ice, the glaciated areas were covered by glossopteris, a distinctive genus of seed fern found nowhere else. The uniqueness and strange distribution of tillites and the glossopteris flora led Du Toit to postulate that all areas bearing them had been grouped together in one continental mass, Gondwanaland, which was close to the South Pole in the Permocarboniferous (see Figure 4b). Other evidences for Gondwanaland include fossils of Mesosaurus, a small reptile found in marine beds immediately above the tillite in South

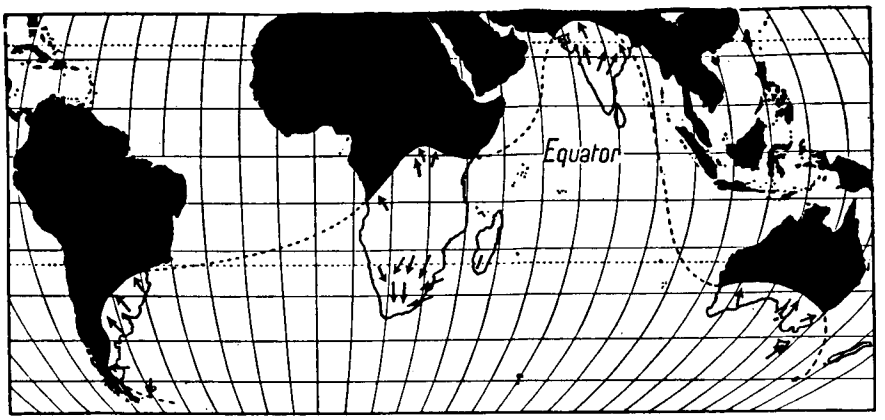


Figure 4a. Map showing the distribution of the late Carboniferous glaciations of Gondwanaland with the continents in their present positions; arrows indicate directions of ice flow (from Holmes, 1965, p. 736).

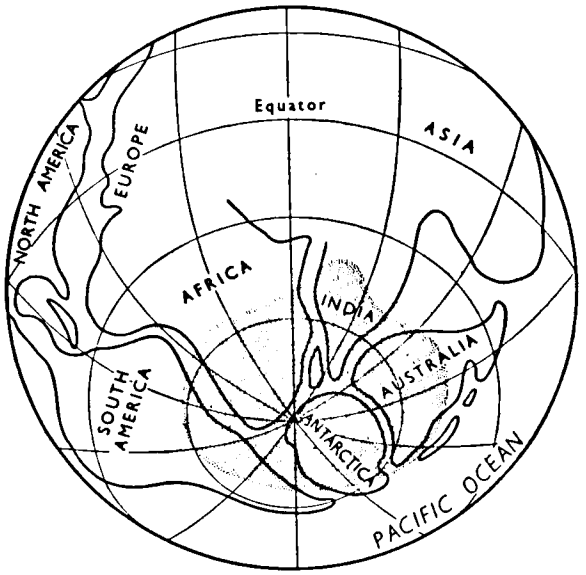


Figure 4b. Reassembly of Gondwanaland proposed by Holmes (1965, p. 736) with southern continents grouped around the South Pole.

America and South Africa, the presence of foreign boulders in the tillites of the Falkland Islands, and the distinctive similarity in each of these separate continents of the whole Permocarboniferous section, which includes everywhere thousands of feet of continental sediments and basaltic lava flows. Without doubt, five continental regions and two islands now widely separated by oceans had a strikingly similar history from the Carboniferous through the Permian.

Other attempts to map the distribution of late Paleozoic climates have been made on the basis of deserts, evaporites, red beds, coal, and coral reefs (see Nairn, 1963). The evidence in every case is ambiguous enough to cause differences of opinion among specialists. One problem arises from the fact that past climates have, on the average, been milder and less rigorously determined by latitude than at present, and they have, therefore, been more susceptible to local influences: the rise of landmasses, changes in ocean currents, and the distribution of epicontinental seas and mountain ranges.

It is obvious in the light of Pleistocene and recent history of New England, for example, that, without changing latitude, some areas can undergo violent changes in climate. It still remains questionable whether a glacial period demands a worldwide drop in mean temperatures; thus, attempts have been made to interpret the Permocarboniferous glaciations in terms of local conditions: a general emergence of landmasses, or a possible land connection between Antarctica and Australia, which altered the ocean currents and temperature. No explanation seems adequate but that is true of glaciations in general. Jeffreys believed that the tillites should be interpreted as a lesson in meteorology rather than using changes in weather as evidence for continental drift. Other opponents of continental drift point out that while Brazil was undergoing glaciation, Peru was accumulating great thicknesses of warm-water limestones, which indicate a severe localization of climates wherever South America lay in the Permocarboniferous. The overseas distribution of glossopteris is ascribed to winged spores. The foreign boulders on the Falkland Islands have been attributed to iceberg rafting, a well-known

mechanism for transport of materials at least as feasible as rafting continents, and Mesosaurus has been credited with crossing the Atlantic during periods of low water resulting from the large ice caps.

Special problems arise in relation to the distribution of land plants. From the time of their first appearance in the Silurian, land plants were all of uniform character and worldwide distribution (Kummel, 1961). In the late Carboniferous the first distinctive floras appeared. These included glossopteris and three other groups: the European, which occurred in Europe, North Africa, and eastern North America; the Angaran of Central and Northern Asia; and the Cathaysian in China and western North America (Figure 5). The distribution of the Cathaysian flora on both sides of the Pacific, but only part way across North America, is at least as mystifying as that of glossopteris. This problem was almost wholly neglected until 1966, when Melville (1966) proposed an arrangement of several protocontinents, including one called Pacifica made up of China and Western North America, and one called Atlantica consisting of Europe, eastern North America, North Africa, and all of South America above the Amazon River. He also retained Du Toit's Gondwanaland and at least one other block. He believes that by the end of the Cretaceous these masses had fragmented, drifted, and merged with other masses into the present configuration. Melville's postulated seam line in North America runs diagonally southeast from the mouth of the Mackenzie River on the Arctic to the Gulf of Mexico west of the Mississippi delta. Structurally, this seam line is wholly inconspicuous in contrast to that marked by the Himalayas, which are thought by many to result from an Oligocene collision between Asia and India as the latter drifted northward.

The known structure of North America hardly allows for the separate development of two halves that became joined longitudinally in the early Cretaceous. Nevertheless, if the distribution of organisms is to be explained by an appeal to continental drift, Melville's elaborate system is a superior one, as it accounts for both the sudden floral break in North America and

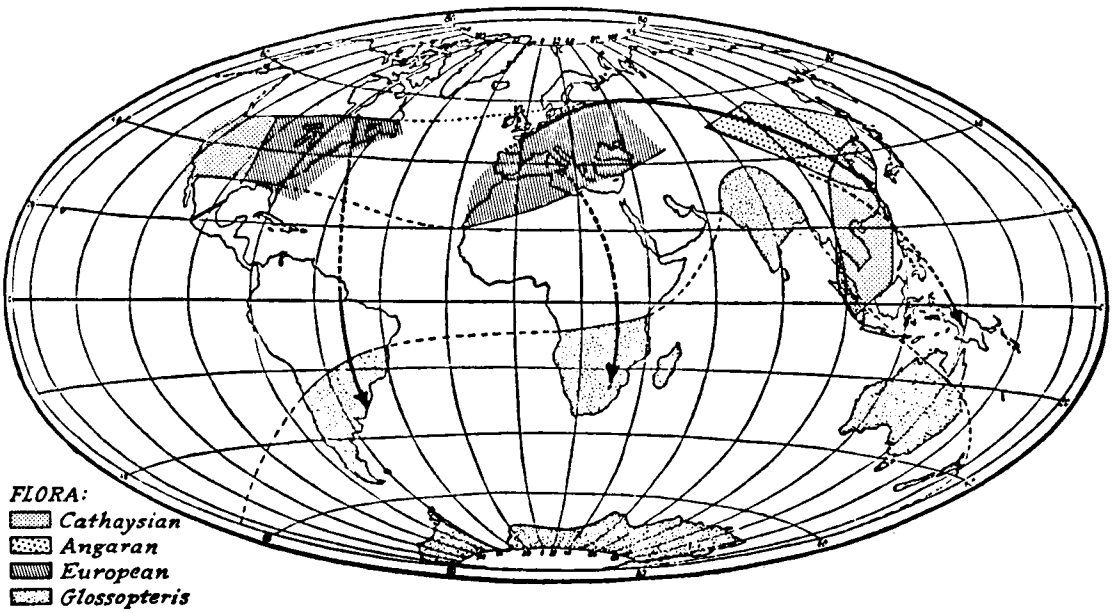


Figure 5. Distribution of late Paleozoic floras. The continents are shown in their present position and size. The three arrows point to the localities where mixed floras of southern and northern elements have been found. Mixed floras have also been reported from Patagonia, Mozambique, and Madagascar (from Kummel, 1961, p. 339).

the identical floras across ocean basins. Like most specific drift systems, however, it proposes solutions to some problems while raising severe difficulties with others.

In summary, the distribution of organisms displays so many anomalies, some of which could not be solved by any configuration of former landmasses, that, although it is widely used, it seems generally unsatisfactory as evidence for continental drift. The strongest evidence still includes that recognized earliest: namely, seven landmasses in the Southern Hemisphere have thick, distinctive Permocarboniferous deposits including the tillites, which are by far the most compelling of paleoclimatic indicators and have not been adequately explained in terms other than continental drift.

2.3 Paleomagnetism

The interpretation of paleomagnetism in rock samples rests upon the following assumptions:

A. Iron-rich oxides of mafic igneous rocks acquire a permanent magnetism oriented toward the earth's magnetic pole, as they cool through their Curie temperature. The iron-bearing detritus and cementing materials in sediments also orient toward the magnetic pole.

B. The magnetic field of the earth is dominantly dipolar and roughly aligned with the axis of rotation today, and this is presumed to have been true throughout planetary history.

C. The earth's axis of rotation cannot change significantly with relation to the ecliptic since the planet behaves like a gyroscope.

Measurements of the fossil magnetism in unweathered rock samples from all continents show that for the past 20 million years, from the Miocene to the present, the locations of the earth's magnetic poles have remained essentially unchanged. Prior to the Miocene, the position of the magnetic pole apparently moved at least 21,000 km over a long curving path from western North America in the late Precambrian, through the present Pacific Ocean to northern Asia in the Mesozoic, and to the Arctic in the mid-Tertiary

(Runcorn, 1963). If rocks of the same age in all continents were to point to the same pole throughout a long migration, this phenomenon could be ascribed to polar wandering. Polar wandering has had two meanings: at one time it implied that the magnetic pole departed radically from the axis of rotation and swiveled independently within the earth; at present, the term is generally understood to mean that, while the magnetic poles remain fixed, an outer shell of the earth, including the crust and part or all of the mantle, becomes decoupled and moves as a unit over the interior — as though the skin of an orange could be loosened and moved over the fruit. Polar wandering of this type would allow the rocks of continents that are stationary, relative to each other, to record an apparent migration of the magnetic poles and of climatic zones. The actual evidence from rock magnetism, however, permits no such simple solution.

Rocks older than the Miocene from different continents point to different pole positions and describe systematically divergent paths of migration (see Figure 6). This suggests that landmasses that were once joined have rifted and moved relative to one another and that their former positions can be reconstructed by superimposing their polar-migration curves. Many investigations have been carried out that purport to show that continental reconstructions based on paleomagnetism coincide well with the requirements of the fit of continental blocks and the pattern of paleoclimates. For example, paleomagnetically determined latitudes of Australia show an apparent migration of over 50° from the tropics in the Devonian toward the South Pole in the Carboniferous. During the same period Australian coral reefs died out and glaciation began (Runcorn, 1963). Correlations of this kind constitute impressive evidence in favor of continental drift.

Criticisms of paleomagnetism are based on the fundamental assumptions, on the rock sampling, and on the interpretations of results. The assumption of a permanently dipolar magnetic field is challenged by MacDonald (1963), who believes the earth may well have been multipolar earlier in its history. If so, movements or rearrangements of pole positions could theoretically be recorded as separate curves from fixed landmasses. Various authorities,

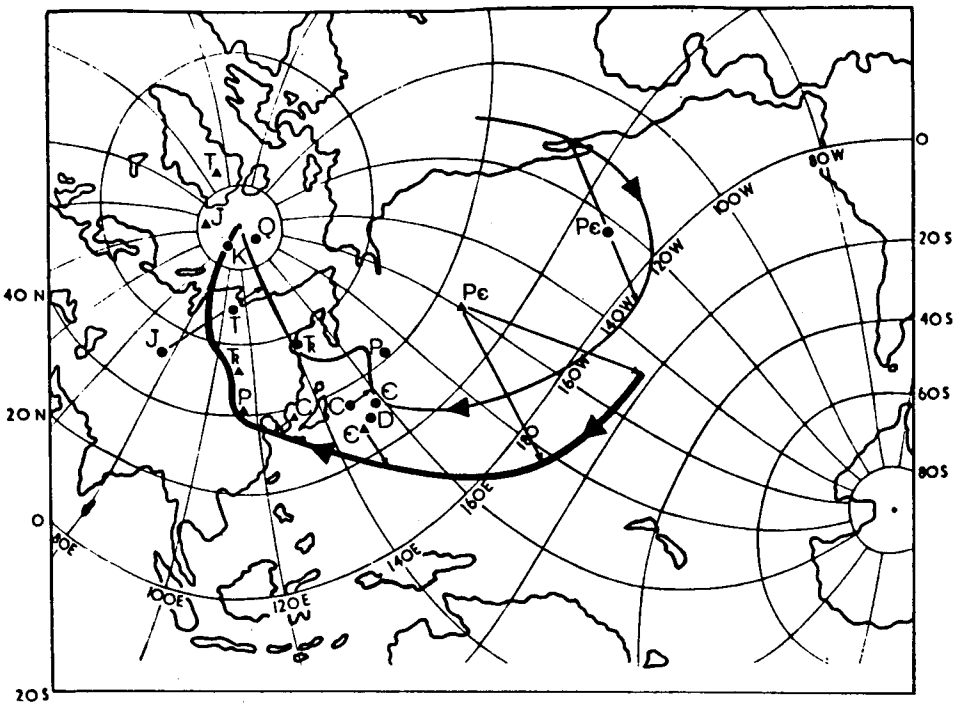


Figure 6a. Polar-wandering paths for Europe and North America derived from the mean poles. ▲ , American rocks; ● , European rocks (from Runcorn, 1963, p. 23).

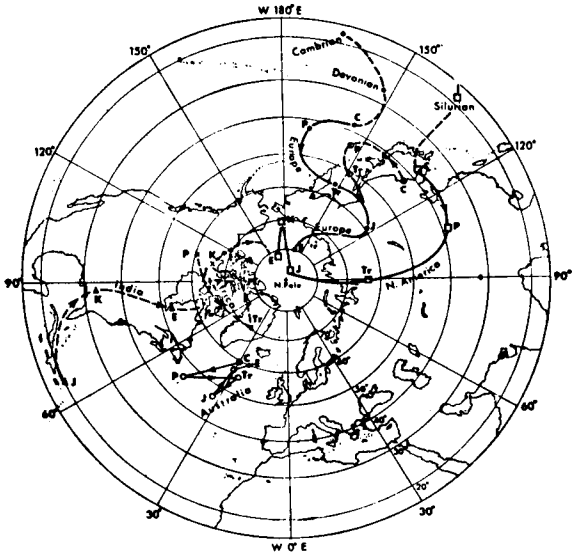


Figure 6b. Polar-wandering paths for India, Australia, Europe, and North America (from Holmes, 1965, p. 1210).

feeling that the sampling has been inadequate, are reserving judgment on paleomagnetic results until suites of samples of known age have been collected from larger areas of the continents. It is well known that pieces of continents move relative to each other, but these authorities feel that the case is less clear for continents as a whole. Paleomagnetic measurements record only the former latitudes of rock samples and give no indication of longitude, a situation that allows much freedom in plotting reconstruction.

Paleomagnetism is a rapidly developing branch of earth science. Every year more measurements are made and duplicated by different investigators. The results are increasingly reliable and appear to be far more solidly based than the objections to the method cited above. On balance, paleomagnetism confronts earth scientists with an impressive body of evidence that, to date, lacks an explanation other than continental drift.

Repeated reversals of the north and south magnetic poles are another phenomenon recorded by fossil-rock magnetism. Pole reversals are world-wide in effect and are, therefore, attributable to an alternation in polarity of the geomagnetic field. So-called self-reversals also occur, owing to special physical or chemical factors in rock units of limited extent.

Pole reversals are not ascribed to the moving of rock masses north and south of the equator and would, therefore, seem to have no relevance to continental drift. They do, however, constitute evidence of fundamental importance to the new concept of ocean-floor spreading, which seeks to explain the youth, topography, and dynamics of the ocean floor, and is viewed by an increasing number of investigators as the mechanism by which continental drift is initiated and maintained (Hess, 1959, 1962; Vine, 1966; Dietz, 1961).

2.4 The Concept of Ocean-Floor Spreading

Recent discoveries in oceanography include the young ages of bedrock, the thinness of sediment, and the complexity of submarine topography dominated by seamounts, guyots, trenches, and midocean ridges.

The oceanic ridges have a maximum relief of ~4800 m and, with all branches, form a continuous system 60,000 km long that bisects the Atlantic, encircles the southern continents, and projects inland along the East African Rift system and the San Andreas fault zone in California (see Figure 7). With local exceptions, the ridges are characterized by shallow earthquakes, volcanism, high heat flow, tensional rift valleys, transverse faulting, and belts of magnetic anomalies. The whole system is young, but some segments are tectonically more active than others.

The pattern of the ridges appears to negate the classical conception of continental drift, because several continents — Africa, South America, Antarctica — have ridges ahead of them as well as in the wake of their supposed paths of drifting. The concept of ocean-floor spreading, however, changes the emphasis from the continents to the ocean floors as the units in active motion.

This concept pictures the ocean floors not as part of the earth's crust but as a thin rind on the upper mantle derived from the latter mainly by a low temperature phase change from eclogite to basalt or peridotite to serpentine. This hypothesis invokes the slow creep of crystalline materials in the upper mantle, with new rock rising up the ridges and spreading away horizontally on both sides. The estimated rates of movement range from about 0.5 to 5 cm year^{-1}. Downward flow, to complete a cycle, may occur along the edges of continents, which are nearly all bordered by deep trenches with or without a full load of sediments, or may occur beneath the continents. The location of descending currents is always more controversial than is that of rising ones. By this mechanism the bedrock of the ocean floors would be constantly renewed by its spreading away from the ridges, as on a system of opposed conveyor belts. As a result, the bedrock of the ocean floor would indeed be young and its cover of sediments thin.

If movement of this type occurs in the upper mantle, sialic continental blocks will presumably be floated passively away from the flanks of active ridges and centered over or between descending currents. If a ridge is

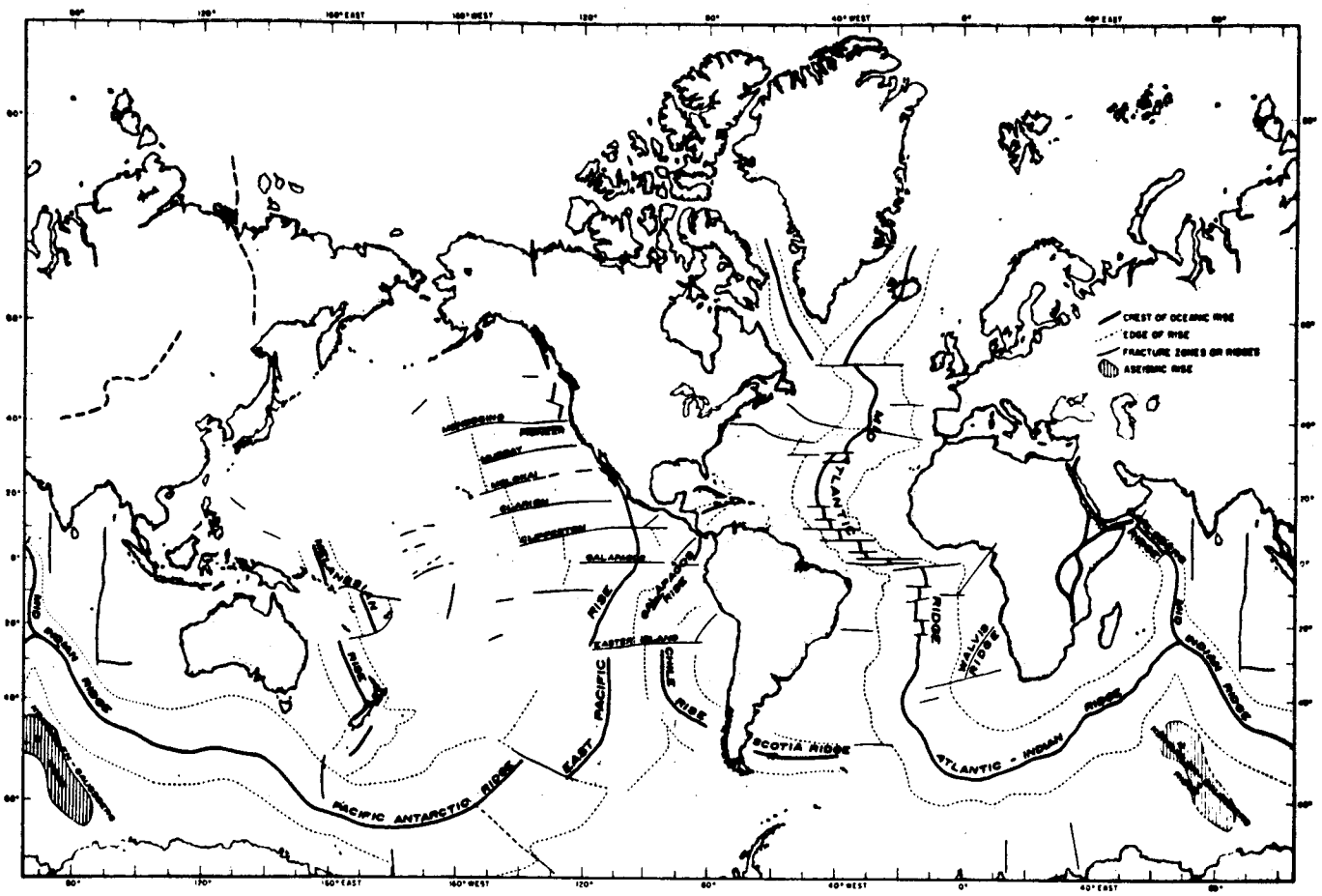


Figure 7. The worldwide system of rises and ridges with related features (from Menard, 1965a, p. 110).

initiated beneath a continent, the eventual result is visualized as a linear uplift followed by tensional necking and rifting and, finally, the symmetrical drifting apart of opposite continental blocks. Something like this may have caused the development of the Atlantic Ocean beginning in the Mesozoic and may be in initial stages today along the East African and San Andreas Rift zones. An alternative view of the latter zone is that the North American continent, in its westward drifting, has partially overridden the East Pacific Rise, which was initiated in an ocean basin (Menard, 1961; Vine, 1966). This circumstance might help to explain some of the features of the Cordilleran region, including the uplift, tensional faulting, and belts of serpentinization.

Ocean-floor spreading has been credited not only with continental drift but also with continental accretion. Hess (1962) and Deitz (1963) have suggested that, as the continents move, segments of the upper mantle, 100 to 200 km deep, move along with them, thus preserving the deep structure of the continents described by MacDonald (1963). (MacDonald, however, would demand a depth of drifting closer to 500 km.) They also postulate that, as the ocean floors approach the continents, they are partially rolled downward and transformed by a second phase change from serpentine back to peridotite and partially crushed against the barrier until a prism of shelf and oceanic sediments and volcanics is welded to the continental margin by orogenic processes. Thus, the continents could show marginal belts of successively younger materials of apparent geosynclinal origin. The folds and thrusts of these belts, however — far from recording crustal shortening, as is generally assumed — would occur in formations that had been added to the continents. Ocean-floor spreading has presumably occurred in different parts of the world during various geological periods, and oceans such as the Atlantic may have opened and closed more than once (Wilson, 1966).

The following observations, among others, are cited in support of ocean-floor spreading.

A. Islands, seamounts, and sedimentary cores reportedly become progressively older from the mid-Atlantic ridge toward the continents. Samples from close to the ridge range from 1 to 18 million years, whereas

the oldest samples, 135 to 150 million years, lie close to both shores. The rates of spreading indicated for each side of the ridge range from 1 to 3 cm year^{-1} (Wilson, 1965a). Several SE-NW-trending volcanic island chains in the Pacific show an increase in age from active volcanoes in the southeast to dormant, eroded, and sometimes sunken islands and seamounts in the northwest. Along the Hawaiian chain, for example, rock ages range from Miocene (20 million years) on Oahu to Cretaceous (100 million years) in the Emperor seamounts, 3000 km to the NW. This is attributed to spreading of the Pacific floor at the rate of up to 3 cm year^{-1} (Wilson, 1963).

The reliability of some of Wilson's data and his interpretations have been disputed by Menard (1965a).

B. Long, linear belts bearing magnetic anomalies that are interpreted as recording reversals of the earth's magnetic poles have recently been found paralleling several oceanic ridges. These belts are symmetrically distributed in alternate strips along either side of the ridges. Continental rocks record five periods of normal (north-seeking) polarity alternating with five of reversed polarity within the past 4 million years. Oceanic rocks should respond to the same reversals and, if it is assumed that these rocks acquire their magnetism as they cool on the ridges, the widths of the anomaly belts indicate how far oceanic rocks have spread laterally during the past 1 to 4 million years and, by extrapolation, over much longer periods (see Figure 8).

The rates of spreading from each side of several active ridges, as calculated from the pattern of anomalies, are as follows (Vine, 1966):

The Reykjanes Ridge, south of Iceland	0.95 cm year^{-1}
The mid-Atlantic Ridge (latitude 38° S)	1.5
The Carlsberg Ridge (latitude 5° N, Indian Ocean)	1.5
The Red Sea floor	1.0
The Juan de Fuca Ridge, south of Vancouver Island	2.9
The East Pacific Rise (latitude 51° S)	4.4

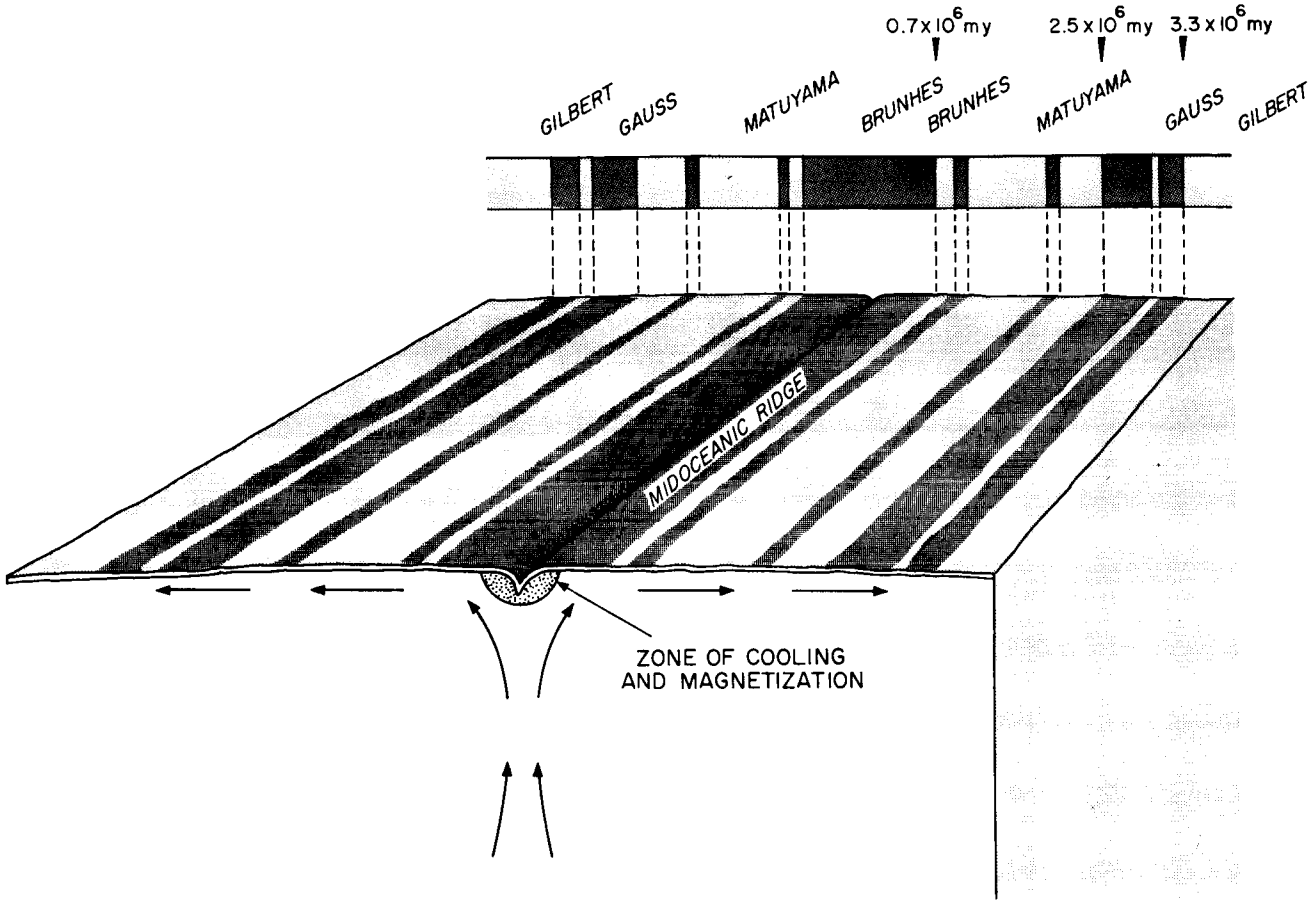


Figure 8. Diagramatic representation of ocean-floor spreading and the "freezing-in" of belts of magnetic anomalies recording reversals of the earth's magnetic poles during the past 4 million years. Names and dates refer to epochs of normal and reverse polarity as measured on continents and islands. Normal polarity (dark); reverse polarity (light). (After Cox, Dalrymple, and Doell, 1967.)

Assuming these rates of spreading, extrapolations of oceanic history have been attempted, not only for the past 4 million years but for the past 11 million years and for much longer periods. Indeed, Vine (1966) suggests that the entire history of the ocean basins, in terms of ocean-floor spreading, is frozen into the oceanic crust. His present interpretations of early Tertiary history, however, are not consistent with those made on the basis of heat-flow data in the Atlantic and Pacific.

Heat-flow values on the East Pacific Rise indicate that this is indeed an area of active upwelling and spreading, but at a rate of only about 1 cm year^{-1}. On the mid-Atlantic Ridge, however, the rate of heat flow is so low, excepting for a narrow zone along the crest, that continuous sea-floor spreading during the Cenozoic appears to be precluded (Langseth, Le Pichon, and Ewing, 1966). These investigators suggest that the mid-Atlantic Ridge may have generated floor spreading and continental drift during the Jurassic and Cretaceous and then lain dormant until active spreading was renewed within the past 10 million years. This interpretation would allow for the formation of the Atlantic Ocean by spreading at the rate of 2 cm year^{-1} (and continental separation at 4 cm year^{-1}) for the first 100 million years, followed by cooling of the ridge until the recent rejuvenation, which accounts for the high heat flow along the axis and the bands of magnetic anomalies recording pole reversals over the past 4 to 11 million years.

In the 7 years since ocean-floor spreading was proposed by Hess (1959), most of its supporters have seen in it the mechanism for causing continental drift. Others have maintained that it occurs without drift by one of the following means:

A. Given an expanding earth, ocean-floor spreading widens the oceans around stationary continents.

B. With no change in the earth's radius, the widening of the ocean floors is compensated for by compressional folding and faulting within the continents.

If the midocean ridges are tensional features along which new material has been added for the past 100 to 200 million years, the proposition that they result from expansion of the earth appears attractive. It has recently been abandoned, however, by one of its main advocates (Heezen, 1966), because it would require a radial expansion of 4 to 8 mm year^{-1} for the past 200 million years to account for the Atlantic Ocean alone. The growth rings of Devonian corals show that the length of the year has increased only 20% in the past 400 million years, and so the maximum possible expansion of the earth in that time has been less than 0.6 mm year^{-1} (Wells, 1963).

The suggestion that the widening oceans are compensated for by crustal shortening in the continents is impossible to demonstrate for the pre-Tertiary, when the continents were presumably in one mass or close together. During the Tertiary an estimated 120×10^6 km^2 of area was added to the ocean floors (Heezen, 1966), and it seems doubtful indeed if an equal area was subtracted from the continents by the crumpling of the Tertiary mountain ranges, including the Andes, Rockies, Alps, and Himalayas. The continents include major tensional features of the same age; folds and thrust faults are notably unreliable indicators of crustal shortening; and, as noted above (p. 56) geosynclinal belts may, in fact, contribute to continental growth.

The conclusion seems most reasonable that, if ocean-floor spreading occurs, the continents are slowly moved over the earth's surface in response to changes in the strength and distribution of convection currents in the upper mantle and in its surface expression, the ocean floors.

2.5 The Convection Problem

All variations of the continental-drift theory depend for their driving force upon convection in the earth's mantle. The most powerful and persuasive arguments against continental drift have always come from those geophysicists who are convinced on both theoretical and observational grounds that convection currents do not occur in the mantle. Their arguments include the following:

A. The upper mantle is too inhomogeneous for convection. Rotating convection cells should cause large-scale mixing, but heat-flow and seismic data indicate major differences in mantle structure beneath the continents and oceans, as well as horizontal layering within the upper mantle.

B. The lower mantle is too rigid for convection. The nonequilibrium figure of the earth, shown by satellite observations, demands a strong, nonplastic earth shell. This requirement cannot be met by the upper mantle, which yields readily to pressures such as the continental ice caps and has a calculated viscosity of 10^{21} poises. It must therefore be met by the lower mantle, which, on this basis, has a calculated viscosity of 10^{26} poises — enough to prevent convective flow at the same temperatures indicated by heat-flow measurements (MacDonald, 1966).

C. Mantle-wide convection, from the core to the crust, would be improbable, even if either the upper or the lower mantle could convect, because the currents would not be sustained across the major transition zone separating the upper and lower mantles.

These objections are aimed largely at the classical conception of one or more pairs of convection cells rotating in a viscous mantle. This conception has a long history and many influential adherents, including Griggs (1939) and Vening Meinesz (1962). If, however, the objections are valid and classical convection cells are not present in the mantle, crustal tectonics result from some other type of mantle motion. Several variations on the convection theory have recently been proposed.

Orowan (1964) points out that, as the mantle is a crystalline solid, it is not subject to Newtonian viscous flow, and therefore MacDonald's calculated strength of 10^{26} Newtonian poises does not apply. He maintains that, like any other crystalline material, the mantle is plastic at moderate temperatures and displays Andradean viscosity (hot creep with no strain hardening) at high temperatures. In such a mantle, deformation tends to concentrate in relatively thin layers, and thus the rising materials take the long, narrow, dike-like forms of the oceanic ridges. Orowan calculates that the Atlantic

floor spreads horizontally from the ridge carrying the Americas westward at the rate of 1 cm year^{-1}. He assumes that there is no downward return flow until the shallow westward-moving layer is underthrust by the eastward-spreading limb of the East Pacific Rise along the Gutenberg-Richter shear system at the Pacific margin of the continents. The mean annual seismic-energy release due to compression along this zone confirms his estimate that the Americas are overriding the Pacific floor at 1 cm year^{-1}. Given movement as slow as this, he concludes that no complete overturn of mantle rock, as in a convection cell, has been accomplished since the Mesozoic. Some investigators speculate that minerals migrating at this slow rate may pass through phase changes en route and thus tend to preserve mantle inhomogeneities.

Elsasser (1965) develops a model whereby convection in the upper mantle is anisometric and largely advective. While a small fraction of the mantle moves vertically up or down, the remaining mass moves in long tongues of material very nearly horizontally. His hypothesis appears to fit well with that of Orowan.

Ramberg (1963) suggests that mantle movements are dominantly vertical (radial) in component. In this case the local variations in temperature or pressure in the upper mantle cause rising or sinking by means of either phase changes or actual movement of materials. Light, heated materials, for example, rise through cooler rock until they cool and come to a stop, or form a bulge and spread laterally, or penetrate the crust as intrusives or extrusives. No systematically rotating return flow is contemplated. More specifically, Ramberg proposes that the chemical differentiation of light, continental sial at a given site produces a heavier residue at depth that slowly sinks as the sial rises. The sinking material drags surrounding mantle toward itself, subjecting the sial to horizontal compression from all sides and causing circumferential tension some distance away from the sial. Menard (1965b) believes that this hypothesis explains better than any other the observed distribution of oceanic ridges, which, as he points out, do not bisect most ocean basins but are distributed on arcs of circles surrounding the continental shields.

Serious objections are raised to all proposed patterns of convection, and the fact remains that we have no satisfactory theory of the forces governing the earth's interior.

The restrictions on any dynamic model of the mantle are, today, more stringent than they were in 1924 when Jeffreys pronounced against continental drift without any knowledge of equal heat flow or mantle inhomogeneity. Nevertheless, evidence from both geology and geophysics indicates that massive movements of the ocean floors and continents are occurring, and, therefore, they must be described and measured with or without an adequate explanation of the energy sources motivating them.

2.6 Crustal Movements

Whatever their cause and significance, crustal movements, both vertical and horizontal, occur at rates ranging from 0.001 to 10 cm year^{-1}, averaged over long periods. The calculated rates of some vertical movements are as follows:

The readjustment of Fennoscandia at the site of maximum thickness of the icecap is measured on the basis of varved clays and the rising and tilting of old shorelines (Flint, 1947):

Maximum depression:	770 m
Uplift to date	520 m
Future uplift	250 m
Time since melting of ice	10,000 years
Rate of uplift:	
First 3000 years	5 cm year^{-1}
At present	1 cm year^{-1}

The readjustment of North America is known with less precision because of the vast area and difficult terrain. The outer limit of warping follows a broad arc from the Dakotas, across the southern end of the Great Lakes to Long Island, and northeastward to Nova Scotia and Newfoundland. From 0 at this line, the upwarping to date, measured from old shorelines, increases northward to about 300 m near Hudson Bay (Daly, 1934).

Approximate uplift to date at	Niagara Falls	50 m
	Kingston, Ontario	100 m
	Lake Superior	200 m
	James Bay	300 m
Estimated future uplift at Hudson Bay		250 m
Present rate at Hudson Bay		0.5 cm $year^{-1}$.

Vertical movements in other areas are estimated as follows:

The Colorado Plateau uplifted 1.6 km in 25 million years	0.018 cm $year^{-1}$
The Tibetan Plateau 4.8 km in 20 million years	0.025 cm $year^{-1}$
The Himalayas 8 km in 20 million years	0.04 cm $year^{-1}$
Geosynclinal sinking (average) 30 km in 100 million years	0.03 cm $year^{-1}$
Hess Guyot, sinking 1.7 km in 100 million years	0.0017 cm $year^{-1}$.

Horizontal movements within continents are generally measured in the folds and overthrusts of orogenic belts or the offset of dated horizons along transcurrent (strike-slip) faults. Analogous measurements are now being made of the offset of linear magnetic anomalies on fault zones in the ocean floors. During the past decade a great many very long fault planes have been recognized with apparent offsets, along the strike, of a few hundred to over 1000 km. These faults border the Pacific basin partly on land, partly in the ocean floor. A series of them cuts east to west across the crests of the East Pacific Rise and the mid-Atlantic Ridge, and across the Carlsberg Ridge in the Indian Ocean (see Figure 6). They have also been found in many other areas, including Scotland, Algeria, and Anatolia.

The significance of these faults with respect to continental drift has been hotly debated. Holmes (1965) sees the movement of crustal blocks along these faults as living proof of drift. Menard (1961) maintains that transcurrent faults would tear a continent to pieces and that, if they had been important in the movement of landmasses, no subsequent fit of continental margins would be possible.

Until the past year or so there appeared to be no doubt that these were authentic transcurrent faults, and the San Andreas of California was the most famous and carefully studied example. Present geodetic measurements show that the Pacific side of the fault is moving northward relative to the continental side, at the rate of 5 cm $year^{-1}$ (Hill and Dibblee, 1953). Jurassic and younger strata are offset along this zone, and on this basis, estimates have been made that the rate of movement has increased from 0.1 cm $year^{-1}$ in the Jurassic, to 1.6 cm $year^{-1}$ in the Pleistocene, to 5 cm $year^{-1}$ today. For comparison, the estimated rates of movement along some other faults are as follows:

Great Alpine fault, New Zealand	2.5 cm $year^{-1}$
Magna Fossa, Japan	2 to 4 cm $year^{-1}$
Algerian faults	1.5 to 3 cm $year^{-1}$.

Recent investigations, however, prove that the San Andreas is by no means a simple transcurrent fault that has had a dominant strike-slip motion throughout its history. Some movement in both directions is indicated, along with thousands of feet of vertical movement of some of the rocks bounded by this fault and its related cross faults (Oakeshott, 1966).

A reconsideration of transcurrent faults is under way, and one probable outcome will be the discovery that none is simple in structure or in history. Another likelihood is that those in the ocean floors, and possibly others, will prove to be partially or wholly transform faults. Transform faults are conceived of as primary dislocations in the continuity of a feature such as an oceanic ridge. If the ridge-forming process was imperfect, the Atlantic Ridge might have originated as the succession of offset segments seen in Figure 3. If ocean-floor spreading has occurred, the ridge crests will have remained stationary while the rock of the flanks will have migrated eastward and westward with, in some cases, the paradoxical relative movement indicated in Figure 9. Transform faults involve horizontal movement of rigid materials but not on the scale that would be indicated if they were conventional transcurrent faults. They are a natural consequence of ocean-floor spreading.

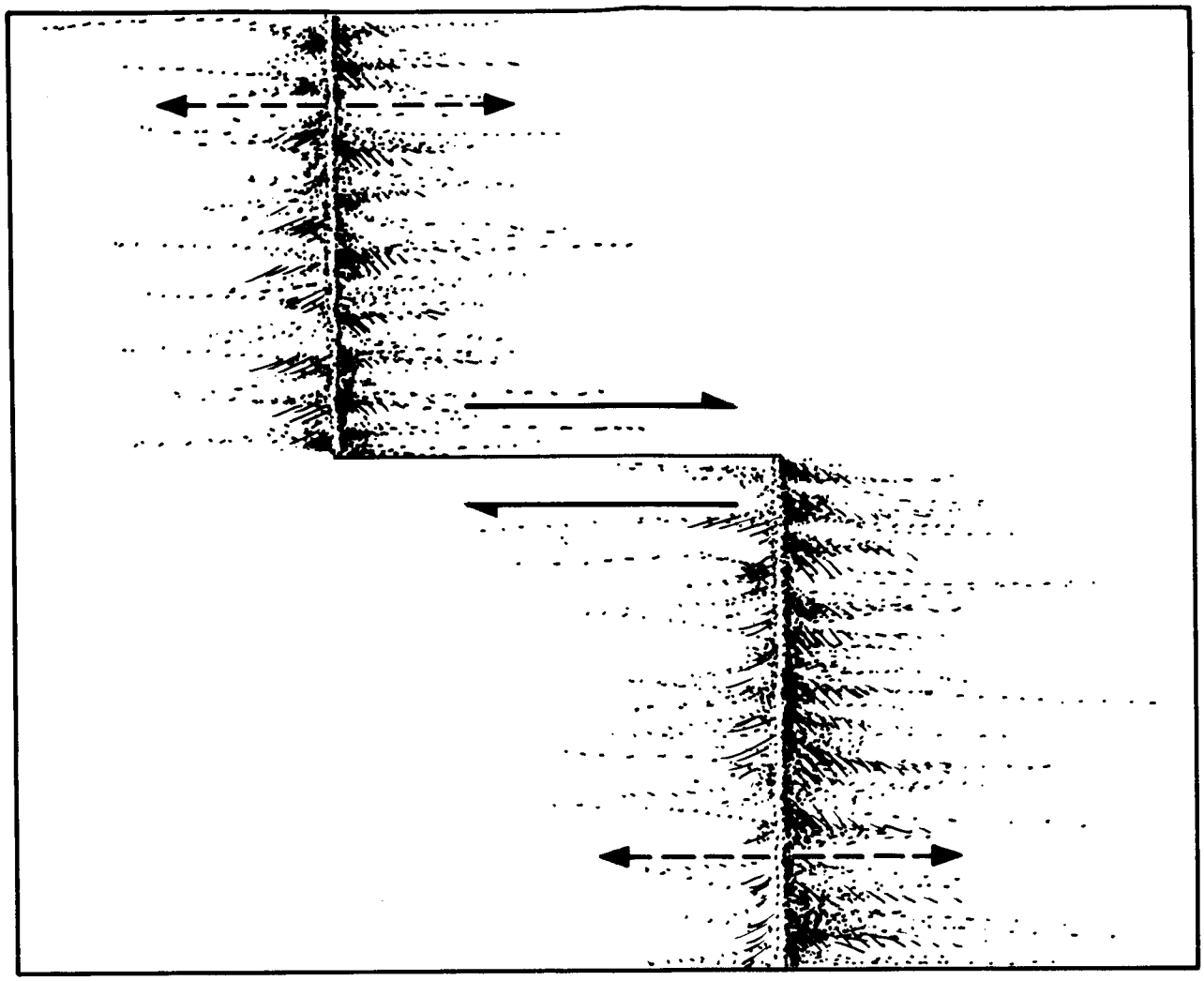


Figure 9. Transform fault connecting two segments of a ridge. Dashed arrows indicate directions of floor-spreading away from the ridge. Solid arrows indicate relative movement along the strike of the fault plane. Note that the movement is the reverse of that along a classical transcurrent fault.

2.7 Estimates of the Rates of Continental Drift

Authorities who believe in continental drift and ocean-floor spreading have cited rates of movement ranging from 1 to 10 cm year^{-1} as follows:

Reference	Location	Rate per year
Orowan (1964)	Americas west from Atlantic Ridge	1 cm
Runcorn (1963)	North America west from Europe since Triassic, 2×10^8 cm 200 million years^{-1}	1 cm
	India north since Jurassic, 6×10^8 cm 150 million years^{-1}	4 cm
	Australia north from pole since Permian, 6×10^8 cm 250 million years^{-1}	2.4 cm
Markowitz, Stoyko, and Fedorov (1964)	Americas west from Europe and Africa from Cretaceous 5×10^6 m 100 million years^{-1}	5 cm
Kropotkin (1965)	Eurasia (Leningrad) approaching California (Ukiah)	10 cm

2.8 Estimates of the Rates of Movement of Ocean Floors

Reference	Location	Rate per year
Wilson (1965a)	Transport of Atlantic Islands from Ridge	3 cm
Wilson (1963)	Elongation of Hawaiian chain	3 cm
Vine (1966)	Separation of belts bearing magnetic anomalies on either side of ridges:	
	Reykjanes Ridge, south of Iceland	1.9 cm
	The Red Sea	2.0 cm
	South Atlantic (latitude 38° S)	3.0 cm
	Indian Ocean (latitude 5° N)	
	Juan de Fuca Ridge, south of Vancouver Island	5.8 cm
	East Pacific Rise	8.8 cm
Langseth, Le Pichon, and Ewing (1966)	Separation across the East Pacific Rise	2.0 cm

3. SUMMARY AND CONCLUSIONS

Of the many lines of evidence offered in support of continental drift, the most compelling in themselves and the most difficult to explain in the absence of drift are:

A. The divergent paths of continental migration, as indicated by paleomagnetism.

B. The symmetrical distribution of belts of ocean floor, showing evidence of alternate pole reversals along either side of the oceanic ridges.

C. The matching of truncated contacts between dated-rock provinces in West Africa and Brazil.

Evidence inconclusive in itself but favorable to drift in combination with the evidence cited above follows:

A. The close fit of continental margins.

B. The distribution of paleoclimates, as indicated by the Permocarboniferous tillites and numerous other organic and inorganic deposits.

Evidence unfavorable to continental drift is the inhomogeneous nature of the upper mantle, as shown by heat-flow and seismic data.

The weight of evidence, much of which has been developed within the past 10 years, indeed within the past 2 years, favors continental drift. The general concept has changed, within the same period, from one of moving continents through yielding ocean floors to that of moving ocean floors and, incidentally, moving continents and possibly related segments of upper mantle along with them.

The apparent rates of the present motion of some continental blocks away from each other range from 1 to 10 cm year^{-1}. In zones of relatively rapid movement it may be possible to verify continental drift by means of laser measurements on satellites, within a few years or, at most, a few decades. It is hoped that laser techniques will soon reduce the error in station positioning to 30 cm. Given this degree of accuracy, a network of at least three laser stations on each of two landmasses could detect a mean shift in relative positions after 5 years of measurements if the landmasses are separating at a rate of 8 cm year^{-1}; after 12 years, if they are separating at 5 cm year^{-1}; and 30 years, if they are separating at 2 cm year^{-1}. Interference from local crustal movements should be detected and corrected for, and this operation would itself constitute important research.

Promising sites for the testing of drift include the relatively active East Pacific, where islands on either side of the rise may be retreating from one another at a rate of up to 8 cm year^{-1}, and those on the east flank of the rise may be approaching the westward-drifting Americas at 5 to 6 cm year^{-1}. The spreading of the South Atlantic could be detected from stations in South America and Africa, if it is indeed occurring at the rate of 3 cm year^{-1}. The Red Sea is a zone where a record of spreading would be of great interest geologically, even though it is assumed to be slow, 1 to 2 cm year^{-1}, and thus could not be determined for several decades.

Within the near future the refinement of laser techniques in satellite geodesy will make it possible, for the first time, to test continental drift by accurate measurements on a worldwide scale. The resulting data on crustal tectonics will be of incalculable value whether or not they confirm the present reality of continental drift.

4. REFERENCES

ANDERSON, D. L.

1965. Recent evidence concerning the structure and composition of the earth's mantle. In Physics and Chemistry of the Earth, vol. 6, ed. by L. H. Ahrens, F. Press, S. K. Runcorn, and H. C. Urey, Pergamin Press, Oxford, England, pp. 1-131.

BULLARD, E., EVERETT, J. E., AND SMITH, A. G.

1965. The fit of the continents around the Atlantic. In A Symposium on Continental Drift, Phil. Trans. Roy. Soc. London, pp. 41-51.

COX, A., DALRYMPLE, G., AND DOELL, R. R.

1967. Reversals of the earth's magnetic field. Scientific Amer., vol. 216, pp. 44-54.

DALY, R. A.

1934. The Changing World of the Ice age. Yale University Press, New Haven, Conn., 273 pp.

DIETZ, R. S.

1961. Continent and ocean-basin evolution by spreading of the sea floor. Nature, vol. 190, pp. 854-885.

1963. Ocean-basin evolution by sea-floor spreading. In Continental Drift, ed. by S. K. Runcorn, Internatl. Geophys. Ser. No. 3, Academic Press, New York, pp. 289-298.

DONN, W. L., DONN, B. D., AND VALENTINE, W. G.

1965. On the early history of the earth. Bull. Geol. Soc. Amer., vol. 76, pp. 287-306.

DOTT, R. H., JR.

1961. Squantum "tillite," Massachusetts — evidence of glaciation or subaqueous mass movement? Bull. Geol. Soc. Amer., vol. 72, pp. 1289-1305.

ELSASSER, W. M.

1966. Thermal structure of the upper mantle and convection. In Advances in Earth Science, ed. by P. M. Hurley, MIT Press, Cambridge, Mass., pp. 461-502.

ENGEL, A. E. J.

1963. Geologic evolution of North America. Science, vol. 140, pp. 143-152.

FLINT, W. F.

1947. Glacial Geology and the Pleistocene Epoch. John Wiley and Sons, New York, 589 pp.

GASKELL, T. F.

1963. Comparisons of Pacific and Atlantic Ocean floors in relation to ideas of continental displacement. In Continental Drift, ed. by S. K. Runcorn, Internatl. Geophys. Ser. No. 3, Academic Press, New York, pp. 299-307.

GRIGGS, D.

1939. A theory of mountain building. Amer. Journ. Sci., vol. 237, pp. 611-650.

HEEZEN, B. C.

1963. The deep-sea floor. In Continental Drift, ed. by S. K. Runcorn, Internatl. Geophys. Ser. No. 3, Academic Press, New York, pp. 235-288.

1966. Lecture at Conference What's New on Earth, Rutgers University, New Brunswick, New Jersey, October.

HESS, H. H.

1959. The AMSOC hole to the earth's mantle. Trans. Amer. Geophys. Union, vol. 40, pp. 340-345.

1962. History of ocean basins. In Petrologic Studies: a volume to honor A. F. Buddington, ed. by A. E. J. Engel, E. L. James, and B. L. Leonard, Geol. Soc. Amer., 660 pp.

HILL, M. L., AND DIBBLEE, T. W., JR.

1953. San Andress, Gavlock, and Big Pine Faults, California. Bull. Geol. Soc. Amer., vol. 64, pp. 443-458.

HOLMES, A.

1965. Principles of Physical Geology. 2nd ed., The Ronald Press Co., New York, 1288 pp.

HURLEY, P. M., MELCHER, G. C , RAND, J. R., FAIRBAIN, H. W., AND PINSON, W. H.

1966. Rb-Sr whole-rock analyses in northern Brazil correlated with ages in West Africa (abstract). Geol. Soc. Amer., Program Ann. Meet., pp. 100-101.

JEFFREYS, H.

1924. The Earth, its Origin, History, and Physical Constitution. 1st ed., Cambridge Univ. Press, Cambridge, England, 429 pp.

KENNEDY, G. C.

1959. The origin of continents, mountain ranges and ocean basins. Amer. Scientist, vol. 47, pp. 491-504.

KNOPOFF, L.

1964. The convection current hypothesis. Rev. Geophys., vol. 2, pp. 89-122.

KROPOTKIN, P. N.

1965. Comments to Professor Bernal. In A Symposium on Continental Drift, Phil. Trans. Roy. Soc. London, pp. 316-317.

KUMMEL, B.

1961. History of The Earth: An Introduction to Historical Geology. W. H. Freeman and Co., San Francisco, 610 pp.

LANGSETH, M. G., JR., LE PICHON, X., AND EWING, M.

1966. Crustal structure of the mid-ocean ridges. 5. Heat flow through the Atlantic Ocean floor and convection currents. Journ. Geophys. Res., vol. 71, pp. 5321-5355.

MACDONALD, G. J. F.

1963. The deep structure of continents. Rev. Geophys., vol. 1, pp. 587-665.

1966. The figure and long-term mechanical properties of the earth. In Advances in Earth Sciences, ed. by P. M. Hurley, MIT Press, Cambridge, Mass., pp. 199-245.

MARKOWITZ, W., STOYKO, N., AND FEDOROV, E. P.

1964. Longitude and latitude. In Research in Geophysics, vol. 2, ed. by H. Odishaw, MIT Press, Cambridge, Mass., pp. 149-162.

MEINESZ, F. A. VENING

1962. Pattern of convection currents in the earth's mantle. Proc. Koninklijke Nederlandse Acad. van Wetenschappen, B, vol. 65, pp. 131-143.

MELVILLE, R.

1966. Continental drift, Mezozoic continents and the migrations of the angiosperms. Nature, vol. 211, pp. 116-120.

MENARD, H. W.

1961. The East Pacific Rise. Scientific Amer., vol. 205, pp. 52-61.

1965a. The worldwide oceanic rise-ridge system. In A Symposium on Continental Drift, Phil. Trans. Roy. Soc. London, pp. 109-122.

1965b. Sea floor relief and mantle convection. In Physics and Chemistry of the Earth, vol. 6., ed. by L. H Ahrens, F. Press, S. K. Runcorn, and H. C. Urey, Pergamon Press, Oxford, England, pp. 315-364.

NAIRN, A. E. M.

1963. Problems in paleoclimatology. Proc. NATO Paleoclimates Conf., Interscience Publ., New York, 705 pp.

OAKESHOTT, G. B.

1966. Geologic evidence for displacement on the San Andreas fault, California. Geol. Soc. Amer. Program Ann. Meet., pp. 151-152.

OROWAN, E.

1964. Continental drift and the origin of mountains. Science, vol. 146, pp. 1003-1010.

RAMBERG, R.

1963. Experimental study of gravity tectonics by means of centrifuged models. Bull. Geol. Inst. Univ. Uppsala, vol. 62, pp. 1-97.

RUNCORN, S. K.

1963. Paloeomagnetic evidence for continental drift and its geophysical cause. In Continental Drift, ed. by S. K. Runcorn, Internatl. Geophys. Ser. No. 3, Academic Press, New York, pp. 1-40.

SUESS, E.

1914. Das aulitz der erde. Prag F. Zempsky, 3 vols., 1885-1914.

VINE, F. J.

1966. Spreading of the ocean floor. Science, vol. 154, pp. 1405-1415.

WASSERBURG, G. J.

1966. Geochronology and isotopic data bearing on development of the continental crust. In Advances in Earth Science, ed. by P. M. Hurley, MIT Press, Cambridge, Mass., pp. 431-459.

WEGENER, A.

1924. Origins of the continents and the oceans. E. P. Dutton and Co., New York, 212 pp.

WELLS, J. W.

1963. Coral growth and geochronometry. Nature, vol. 197, pp. 948-950.

WILSON, J. T.

1963. Evidence from islands on the spreading of ocean floors. Nature, vol. 197, pp. 536-538.

1965a. Evidence from ocean islands suggesting movement in the earth. In A Symposium on Continental Drift, Phil. Trans. Roy. Soc. London, pp. 145-167.

1965b. Transform faults, oceanic ridges, and magnetic anomalies southwest of Vancouver Island. Science, vol. 150, pp. 482-485.

1966. Lecture at Conference What's New on Earth, Rutgers University, New Brunswick, New Jersey, October.

VAN WATERSCHOOT VAN DER GRACHT, W. A. J. M.

1928. Theory of continental drift. A Symposium Amer. Assoc. Petrol. Geologists, Tulsa, Okla., 240 pp.

SOME GEOPHYSICAL IMPLICATIONS OF THE SATELLITE GEOPOTENTIAL

Chi-Yuen Wang[1]

1. INTRODUCTION

Dr. Lundquist read, during the introductory session, a passage written about 10 years ago by Dr. Whipple, predicting that, as a byproduct of satellite geodesy, there would be a significant improvement in the computation of the density distribution inside the earth. I will first review what has been done in this direction, and then discuss the mechanical characteristics of the earth as inferred from the satellite geopotential. Most of the problems I will mention are still unsolved.

2. DENSITY DISTRIBUTION IN A SPHERICAL EARTH

The discussion of density distribution inside the earth is largely an analysis of the variation of the velocities of the two seismic body waves with depth. Figure 1 shows these velocities plotted against depth.

[1] Physicist, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts.

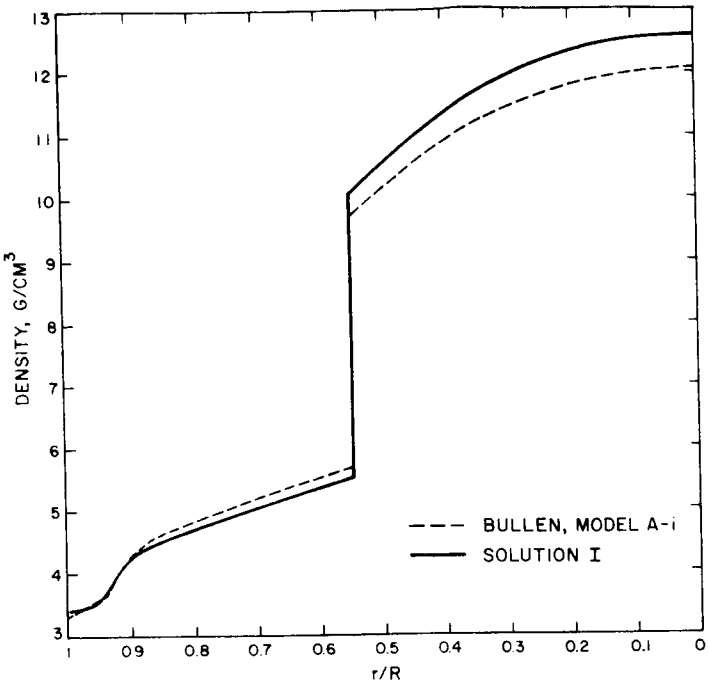


Figure 1. Density versus radius for model A-i and solution I (from Birch, 1964).

Analysis indicates that in layer D the variation is consistent with that expected from a homogeneous layer under compression with an adiabatic thermal gradient. Variation in layer C is too large for such an expectation. The temperature in the upper mantle, or layer B, may be much higher than the adiabatic gradient, but how much higher is not known.

The uppermost few tens of kilometers are commonly termed the crust. This thin but complex layer is an insignificant fraction of the earth's mass, and is conventionally represented in a model for the calculation of the interior density by a crust with uniform thickness and density.

A most fruitful discovery from earlier geophysical investigation was the close association of density change with the rate of change of the velocities of the seismic body waves.

In the last few years, an empirical relation between the variation of density and that of compressional velocity has been established on the evidence of high-pressure studies of rocks and minerals in the laboratories. The relation states that for silicates and oxides of the same iron content the density is roughly proportional to the velocity of compressional waves.

In the computation of density distribution inside the earth, this empirical relation can be applied to the whole mantle with reasonable results. But it is preferable if we apply the Adams-Williamson relation to the lower mantle and the core, where an empirical relation has not yet been established.

If we define the slope or the shape of the density curves by the velocity curves, using the empirical and the Adams-Williamson relations, we still need to specify the initial values of the density curves, that is, the densities beneath the discontinuities. We have two major discontinuities: one at the top of the mantle, and another at the top of the core.

The densities at these discontinuities are determined by postulating two conditions: Both the resultant total mass and the resultant moment of inertia should be equal to the observed values of the earth.

Now, a significant change in our determination of the moment of inertia of the earth has resulted from satellite geodesy. Perturbations of the satellite orbits give a very accurate determination of J_2, the coefficient of the second zonal harmonic in the geopotential. The ratio of J_2 to the dynamical ellipticity gives the ratio of the moment of inertia about the polar axis to the product of the mass of the earth and the square of the equatorial radius. The satellite J_2 gives this ratio as 0.3306, compared with the formerly adopted value 0.3336, which was computed from the flattening 1/297 of the international ellipsoid.

This change in the value of the moment of inertia of the earth implies a decrease of density in the outer part of the earth and a corresponding increase of density in its inner part. In Figure 1, two density distributions

computed for a spherical earth are given. The dashed curve, computed with the old value of the earth's moment of inertia, and the solid curve represent the improved density distribution by use of the value of the satellite J_2.

The resulting densities in the lower mantle are in good agreement with shock-wave data on rocks with FeO contents in the range of 10% by weight.

3. THE ZONAL HARMONICS IN THE EARTH'S SHAPE

By applying the density distribution found from the spherical assumption along the mean radius of the ellipsoidal earth and by assuming that hydrostatic stresses exist at every point in the interior, we can calculate the corresponding shape. This is the shape of a hydrostatic earth having the same ratio of the moment of inertia about its polar axis to the product of the earth's mass and the square of its equatorial radius as that of the actual earth.

It has been found that the ellipticity of the actual earth differs from that of the hydrostatic earth so computed. The actual ellipticity is about 1/298. 3, and the hydrostatic ellipticity, about 1/300. If the difference between the dimensions of these two ellipsoids is described by a second-degree zonal harmonic, the coefficient of this harmonic is 80 m.

What is the physical significance of this difference, and what are its causes? Before answering, let me review two interesting explanations for the nonhydrostatic ellipticities.

Munk and MacDonald (1960) suggested that the differences are due to the delayed adjustment of the rocky earth to deceleration and that the present ellipticity corresponds to the hydrostatic ellipticity of 10 million years ago. To reconcile this length of time with the information derived from the Fennoscandian uplift, Takeuchi (1963) and Takeuchi and Hasegawa (1964) suggested that the earth contains a rigid lower mantle and a viscous layer in the upper mantle; this implies a long recovery time because the flow is restricted to the upper mantle. Dicke (1966) remarked that a detailed study of gravity anomalies over Canada (Innes, 1960) provides a rough test for this model:

> If this model were correct, the longer waves should have decayed hardly at all, and the 100 mgal gravity anomaly (generated by depression due to ice load) should still be present. An examination of the anomalies over large sections of Canada shows no substantial departure from isostasy except for localized deformations. The longer wave distortions contribute... to the gravity anomaly probably under 10 mgal.

It should also be noted that this model, as proposed by Takeuchi, is incompatible with the great gravity anomalies over the Hawaiian archipelago, the Indonesian Strips, and part of the Green Mountains, as we will discuss later.

Dicke proposed a completely different model. He suggested that the temperature under the equator is slightly greater than that under the poles and that the mantle responds to the stresses by flowing instead of by supporting them rigidly. There is little evidence from the surface heat flow that the temperature under the equator is any greater than that under the poles. Furthermore, a higher temperature under the equator might have just the opposite effect: It might cause greater thermal expansion and therefore smaller density.

Daly (1934) showed that the maximum extent of the Pleistocene icecap reached 46° N over Europe and 40° N over North America. The surface of the land from the North Pole to latitude 45° N has an area of about a quarter of the total land area of the earth. If we assume that this surface was covered with ice an average 3 km thick, then the overall sea level at that time should have been more than 200 m lower than it is now. The deformation of the rocky earth must be complex under the redistributed load, partly owing to the irregular distribution of the load.

If a huge amount of ice is piled up on both polar caps for a long time, the figure of the solid earth may become flattened. The ice started to melt rapidly some 11,000 years ago. After a large part of the ice has melted and before sufficient time for complete isostatic adjustment has passed, the water

levels at high latitudes should be higher than those before the ice age, and those at low latitudes should be lower than before. If this explanation is correct, a considerable portion of the observed changes in sea levels is of recent origin — perhaps some tens of thousands of years ago.

The justification for this speculation lies in positive answers to the following two questions:

A. Is the rheology of the earth's mantle compatible with this idea?

B. Was the ice load during the Pleistocene large enough to cause such elliptical deformation?

There are still many unanswered questions regarding the rheology of the mantle under prolonged shear stresses. Although Griggs, Turner, and Heard (1960) showed that the behavior of a number of rocks is almost ideally plastic, the existence of finite strength under prolonged stresses cannot be considered as established by their experiment. Experimental data on crystalline material subjected to small stresses for a very long period of time do not exist, but a few theoretical results on idealized material under a differential load may be of interest here. Haskell (1935) demonstrated that, for a semi-infinite viscous medium under a loaded strip, most of the lateral flow takes place below a depth equal to or greater than the width of the load. For an ideally plastic medium under the same load, Prandtl's solution, as modified by Hill (1950), indicates that the gliding surfaces occur directly below the load and that most of the lateral displacement takes place at depths less than half the width of the load.

Applying Haskell's and Prandtl's solutions to areas covered or previously covered with huge icecaps yields distinctly different types of response from the earth. For the viscous model, a load as wide as that of Fennoscandia or Antarctica will affect most of the mantle under the disturbed area. We expect the lateral flow to spread out in the mantle deeply, as well as far from the loaded region. On the other hand, for a plastic model we expect

most of the displaced material to pile up in the periphery of the load. This piling-up of material would manifest itself in a peripherally bulged topography and peripheral anomalies of gravity around the loaded region. These peripheral phenomena would be absent or negligibly small on a viscous model.

The hypothesis that there should be pronounced peripheral phenomena was proposed by Jamieson early in 1865. However, a hundred years of geological surveying has failed to bring out any substantial evidence to support Jamieson's hypothesis. Although gravity anomalies are positive near the island of Bornholm, Denmark, and in the southern part of Norway, they are due to other causes (Heiskanen, 1936). Recent studies on the bathymetric characteristics of the sea bottom around the polar areas (Zhivago, 1962; Theil, 1962; Ostenso, 1962, 1963) have further confirmed the absence of the topographic bulges around any loaded (or previously loaded) areas.

That Antarctica and the continental regions around the Arctic have been depressed under the ice load can be illustrated by some gravitational and topographic features within or around those regions. Gravity surveys on Antarctica (Woollard, 1962; Bentley, 1964; Frolov, 1964) have shown that the continent is approximately in isostatic equilibrium, although a recent lowering of the ice surface by a few hundred meters may be undetected (Bentley, 1964). Deep circumferential fractures around Antarctica were located and interpreted as having been caused by isostatic adjustment of the crust under the load of the icecaps (Zhivago, 1962; Kort, 1964; Frolov, 1964). Similar fracture systems were found around the margin of the Arctic Ocean (Heintz, 1962; Ostenso, 1962, 1963) and the glaciated area in North America (Daly, 1934). The largely isostatic compensation of Antarctica and the circumferential fracture system around Antarctica, the Arctic basin, and the glaciated area in North America clearly indicate that fractures and flows are the types of deformation of the earth under the huge load of ice. Although we have strong evidence that the mantle is weaker from the top to 400 km deep, we do not know whether the mantle stays weaker or gets stronger again below 400 km. The absence of peripheral bulges around any of the

loaded or previously loaded regions suggests, however, that slow lateral flows under huge icecaps take place deep in the mantle.

The linear viscous earth model was studied by Haskell (1935). The application of this model to many geological features has led, however, to some clearly conflicting inferences about their decay rates. The best single example is the Hawaiian archipelago, which consists of a group of volcanic islands of different ages. Although most of the volcanoes were extinguished many millions of years ago, the Hawaiian group is still growing. Its rocks are mostly 1 million years old, but those of Kauai are about 10 million years old (Strange, 1966a); therefore, the islands must differ in age by some millions of years. If viscous flow is the rule for deformation under the loads of these volcanoes, we would expect to see systematic differences in the Bouguer gravity anomalies over these islands, differences that decrease as the ages of the islands increase. Dr. Strange kindly showed me a map of the Bouguer anomaly over the entire Hawaiian archipelago, which shows that no such differences exist. Maximum anomalies of about 300 mgal and average anomalies of about 250 mgal appear to exist in all the volcanic islands. It seems that through millions of years these islands have been standing still. Either there is no isostatic adjustment in this region or the decay time is far too great for sinkings to be detected. A similar example is the Mio-Pliocene Lake Nevada (Lovejoy, 1964); although the water level decreased a few hundred meters between the Miocene and Pleistocene epochs, no evidence of isostatic rebound was found. Other famous examples include the great anomalies over the Indonesian Strips (Vening Meinesz, 1934) and over part of the Green Mountains (Diment, 1954). These evidences indicate that for surface loads of small dimensions (less than a few hundred kilometers in width) the stresses relax at an extremely low rate. A case that negates the above inference is the uplift of the ground at Lake Bonneville after its effluviation (Crittenden, 1963). The computed viscosity from Lake Bonneville should not, however, be applied to the general character of the earth as a whole.

All that can be said at the moment is similar to the classical lithosphere-asthenosphere hypothesis: the outer layer of the earth is strong, capable of withstanding stress differences up to a few hundred bars over some millions of years. Below this layer, the deeper mantle is much weaker. Under surface loads of small dimensions, stress distributions are not affected beyond the strong, outer shell; the result is either an elastic deformation or an inelastic one with an extremely long decay time. On the other hand, under surface loads of large dimensions, such as those of Fennoscandia and Antarctica, stress differences of considerable magnitude penetrate deeper into the weaker part of the mantle and, together with the bending of the outer shell, cause lateral flow of the mantle away from the regions under the load. This induces instability in the outer shell and results in peripheral fractures around the loaded areas. Thereafter, the blocks on the two sides of the fracture are held only by friction and some degree of welding where the temperature is high enough to cause recrystallization.

Let us suppose that before the ice starts melting the loading time is long enough for the viscous earth to attain isostasy. The displacement of the rocky surface corresponding to a lateral flow of mantle material of average density 5 is $-0.2 \times P_2$ km.

Rapid retreat of the ice sheets began 11,000 to 15,000 years ago. Assuming an exponential decay time of 50,000 years, the surface deformation should now have decayed to $-80\ P_3(\cos\theta)$ m; this is the difference between our best-fit and the hydrostatic ellipsoids. At the same time, the decay period of 50,000 years is also adequate for the long-wavelength gravity anomaly over Canada to decay to its present value, probably 10 mgal. We do not attempt to derive this decay time through simple schemes of model study, since the actual behavior of the mantle is unknown and complex. The importance of this decay rate lies in the fact that it fits the picture of the present $n = 2$ distortion of the earth's figure as the residual of the recovery of the deformed earth after the melting of a great amount of polar ice.

The present rate of recovery at the pole is 1 cm $year^{-1}$. If we assume that the recovery of the ellipticity of the earth is accompanied by a flow of material that induces a change of density in the entire mantle described by $d\rho/dt = C P_2 (\theta)$, where C is a constant, the corresponding dw/w of the earth's rotation is found to be 7.84×10^{-11} $year^{-1}$. Dicke (1966) estimated from his study of ancient eclipses that this value was 7.54×10^{-11} $year^{-1}$.

The best we can now say is that the acceleration found by Dicke can be explained by certain configurations of internal flow of mass in response to the recovery of the earth's ellipticity, and that the acceleration does not imply a change of the gravitational constant G. The present argument, however, cannot be used to calculate the amount of acceleration with any accuracy.

Next to the nonhydrostatic J_2, the greatest discrepancy from a hydrostatic earth is represented by J_3, the coefficient of the third-degree zonal harmonic in the earth's potential. It corresponds to a pear-shaped geoid, with its more pointed end at the North Pole. The zonal harmonics in the geopotential have been determined up to the 14th degree (Kozai, 1964). The geoid corresponding to the combination of these 14 zonal harmonics persists in having a pear-shaped appearance, quite similar to that resulting from the third-degree zonal harmonic alone. It therefore appears that the pear-shaping component is one of the principal factors in the earth's shape.

Eardley (1964) found, under the assumption that the rocky earth was rigid, that sea levels at the equator have fallen about 200 m since the Cretaceous period and have risen around the poles by the same amount or more. He suggested that this receding of sea water from the equatorial region and piling-up at the polar regions were due to the deceleration of the earth's rotation since the Cretaceous period. About one-tenth of the observed changes in sea level is then due to elastic displacement of the rocky surface. The deformation of a viscous earth model is much less understood. The displacement of the rocky surface depends critically on the exact nature of the flows. It is safe to say, however, that in order to restore isostasy a

vertical displacement of the rocky surface about twice that for an elastic earth may result, if the average density of the outflowing material is assumed to be about 5 g cm^{-3}, since, as will be shown, the flows probably take place deep in the mantle. Given enough time, a residual in the gravitational potential smaller than that for an elastic earth would appear. The actual deformation of the earth due to the redistribution of sea water must be more complex, partly because the surface load cannot be simply described by a second-degree zonal harmonic and partly because the behavior of the mantle may not follow the rules for a simple elastic or viscous model.

As water flows to attain a new equilibrium position, its redistribution is restrained by land masses over the earth's surface. Around the North Pole, water can flow freely, but it is blocked by the Antarctic continent at the South Pole. About three-fourths of the other continents are in the Northern Hemisphere, whereas over 90% of the total area between 30° S and 60° S is covered with water (Kuenen, 1950). During the redistribution of water, mass has been accumulated over the oceanic areas. Since the zonal harmonics represent the rotational average, the zonal figure of the geoid over the North Pole and most of the Southern Hemisphere should be relatively higher than that over Antarctica and most of the Northern Hemisphere. The result gives a pear-shaped component.

If we assume that the pear component of the geoid arises from a surface density at the surface of the earth, it can be shown that this density is equivalent to a layer of water with a thickness 280 m at the North Pole. Unless the agreement between this result and Eardley's study of sea level is a coincidence, the third-degree zonal harmonic component of the earth's figure must be the combined result of the asymmetrical distribution of land and sea and of the redistribution of sea water.

4. THE TESSERAL HARMONICS IN THE EARTH'S SHAPE

Let us turn to the tesseral harmonics. In recent years, many heat-flow measurements have been made over the earth's surface. The distribution of the measurements, however, is very uneven; most of them are in the oceans and concentrated in the East Pacific Rise, off the east coast of the United States, and off the west coast of Europe. The rest are scattered in North America, the British Isles, Japan, Hungary, Germany, Australia, and the southern tip of Africa. Large gaps exist for certain continental and oceanic regions where no measurements have yet been made.

The first compilation of worldwide heat-flow data appeared in 1963 (Lee, 1963). With these data, I (Wang, 1963) noticed that when both sets of data were averaged into $20° \times 20°$ squares, there were some similarities between the distributions of the heat flow and the geoidal undulation (Izsak, 1963a) in the sense that most of the squares with heat flow higher than 1.5×10^{-6} cal cm^{-2} sec^{-1} correspond to squares with negative geoidal height, while most of the squares with lower heat flow correspond to squares with positive geoidal height. Lee and MacDonald (1963) gave harmonic representations of the heat flow with spherical harmonic coefficients to the second order. They also noticed a negative correlation between the pattern for the distribution of heat flow and the geoidal undulation. In a test of this correlation, this harmonic representation of heat flow was compared with the satellite geopotential. So that both sets of data would have similar wavelengths in their distribution, I chose harmonic coefficients for the geopotential determined only to the second order. (Using Lee and MacDonald's (1963) analysis (Table 2, column 4) with the extreme values deleted, and Izsak's (1963b) coefficients $C_{2,2} = 6.98 \times 10^{-7}$ and $S_{2,2} = -5.74 \times 10^{-7}$, I computed the heat flow and geoidal height for each $10° \times 10°$ square.) At 95% confidence the population correlation coefficient was found to be -0.82 ± 0.02. The

distribution of points for geoidal height versus heat flow, with the line of regression, is shown in Figure 2. However, we need many more heat-flow measurements before any definite statement about the correlation with gravity data can be made.

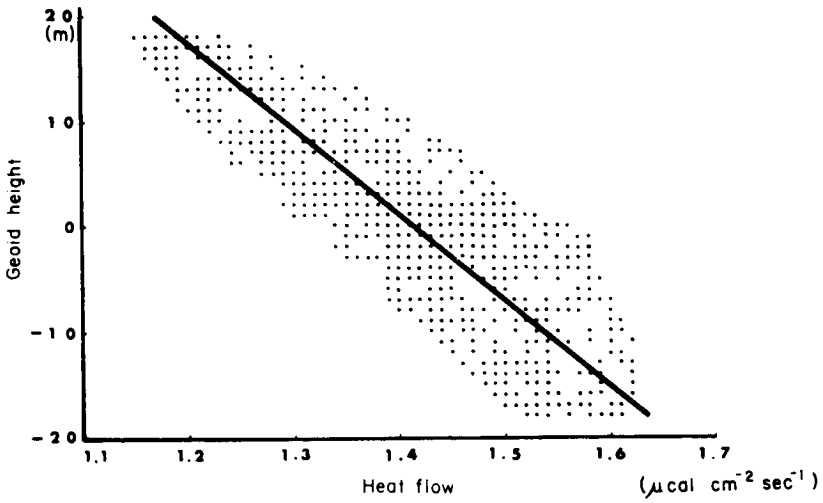


Figure 2. Heat-flow data and the regression line.

Radiometric determinations of uranium, thorium, and potassium have shown that radioactivity of rocks varies through a wide range, even in the same type of rock. In an example given by Lovering and Morgan (1963) on the uranium and thorium abundances in eclogites, uranium varies from 0.018 to 0.24 ppm and thorium from 0.015 to 0.60 ppm; the ratio Th/U is therefore highly variable. It is reasonable to assume that the radioactivity of the upper mantle is not laterally homogeneous. If the temperature perturbation related to a laterally inhomogeneous distribution of radiogenic heat sources causes uneven thermal expansion in the interior, a density anomaly is produced. My purpose in this section is to find a model in which the distributions of temperature and heat sources will give the observed variations in the surface heat flow, and, through thermal expansion, the magnitude of the density anomaly necessary to explain the geoidal undulation.

Since we are only concerned with the deviations of heat flow and geoidal height from some reference levels, it is necessary to define the physical meaning of the last. The reference state is the distribution of thermal properties and mass, which gives the spherical and the elliptical terms in the geopotential and the average heat flow on the earth. The residual variations in both the heat flow and the geopotential are controlled by some factors in the interior that deviate from the reference state.

I will show later that the distributions of temperature and heat sources that control the residual variations in both the heat flow and the geopotential are located in the upper mantle, where the dominant process of heat transfer is thermal conduction.

At temperatures below about 600° C, the mean thermal conductivity K of the rocks likely to compose the outer part of the earth may be taken as 0.006 cal cm^{-1} sec^{-1} deg^{-1} (Birch and Clark, 1940). At higher temperatures, radiative transfer of heat may start to play a certain part in the total conductivity to keep it from falling below this value. At temperatures of the order of 1200 to 1700° C, the contribution from radiative transfer may be equal to or greater than that from the ordinary conductivity (Clark, 1957).

Adopting K = 0.006 cal cm^{-1} sec^{-1} deg^{-1}, the thermal expansion $\alpha = 5 \times 10^{-5}$ deg^{-1} for eclogite (Birch, 1952), and Lee and MacDonald's (1963, Table 2, column 4) heat-flow coefficients, we find that the inhomogeneous distribution of the heat sources is within the outer 100 km of the earth. Figure 3 gives the corresponding inhomogeneous distribution of temperature in the middle of this outer 100-km layer, and Figure 4 gives the variation in the rate of heat production herein.

Increasing thermal conductivity decreases the temperature inhomogeneity. To view this effect, we double the thermal conductivity, i.e., K = 0.012 cal cm^{-1} sec^{-1} deg^{-1}. If we keep the inhomogeneous distribution of heat sources the same, then the temperature and the geopotential are halved. For the geopotential to reach the same magnitude as the residuals in the satellite

geopotential, the thickness of the outer layer where the inhomogeneous distribution of heat sources is located must be increased to 220 km. The distribution of temperature in the middle of this layer is similar to that given in Figure 3, and that of heat production is about half that shown in Figure 4.

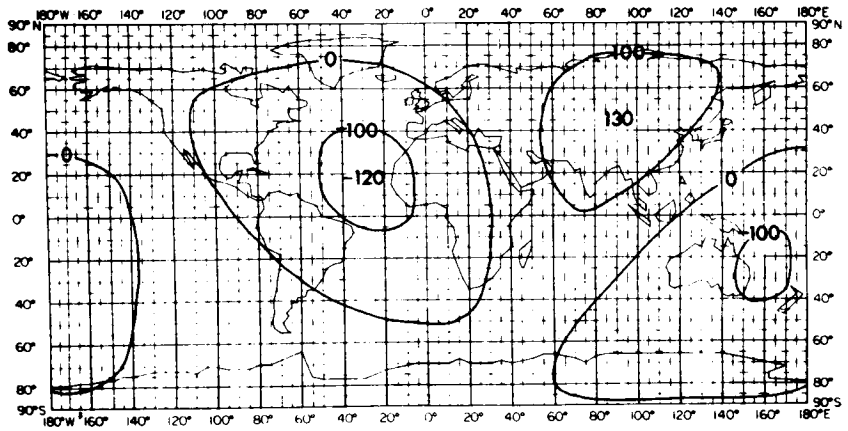


Figure 3. Temperature perturbations with respect to the reference state (isotherms in degrees celsius) at depth 50 km.

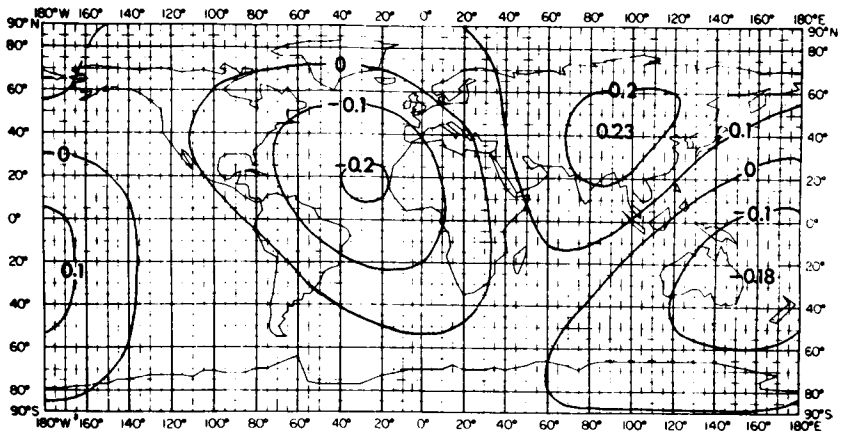


Figure 4. Variations of heat sources A in units of 10^{-13} cal cm^{-3} sec^{-1} with respect to the reference state.

Since the mantle under the continents may be different from that under the oceans, the temperature curves given in Figure 3 do not represent the actual variations of temperature in the mantle. A more realistic picture cannot be given until the temperature distribution in the assumed reference state is found. The temperature curves in Figure 3 imply, however, that isothermal surfaces in the upper mantle cannot be simple geometric surfaces.

Since rocks cannot have negative radioactivity, the above results on the fluctuation of heat sources suggest that, with reasonable choices of the average conductivity, the average rate of heat production in the upper 100 to 200 km of the mantle should be at least 1 to 2×10^{-14} cal cm^{-3} sec^{-1}. In any case, these rates are higher than those of peridotite and dunite but are lower than those of basalt and some eclogites (Lovering and Morgan, 1963; Tilton and Reed, 1963). An eclogitic upper mantle is hence in this respect more favored than one composed of peridotite or dunite.

We still need many more heat-flow measurements before any definite statement about the correlation with the gravity data can be made. As the present distribution of the heat-flow determinations is not adequate for a detailed harmonic analysis, the correlation between these two sets of data can only be regarded as a speculation.

Recently, Dr. William Strange (1966b) remarked that there seemed to be a certain correlation between the geoidal undulations of the SAO Standard Earth and some of the major geologic structures. As most of the major structures have widths shorter than the 8th-degree harmonic, correlation of this sort will be more meaningful if the harmonic expression for the geopotential is expanded to a higher degree.

5. CONCLUSIONS

The geophysical implication of satellite geodesy is that there are large-scale temperature or chemical inhomogeneities within the earth, inhomogeneities that must be supported by stress differences. Stored within the disequilibrium figure is a significant source of energy. How is this energy released? There is no simple answer to this question. Much theoretical and experimental work lies ahead.

6. REFERENCES

BENTLEY, C. R.

1964. The structure of Antarctica and its ice cover. In Research in Geophysics, vol. 2, ed. by H. Odishaw, MIT Press, Cambridge, Mass., pp. 335-389.

BIRCH, F.

1952. Elasticity and constitution of the earth's interior. Journ. Geophys. Res., vol. 57, pp. 227-286.

BIRCH, F., AND CLARK, H.

1940. The thermal conductivity of rocks and its dependence upon temperature and composition. Amer. Journ. Sci., vol. 238, pp. 529-558, 613-635.

CLARK, S. P.

1957. Radiative transfer in the earth's mantle. Trans. Amer. Geophys. Union, vol. 35, pp. 931-938.

CRITTENDEN, M. D., Jr.

1963. Effective viscosity of the earth derived from isostatic loading of Pleistocene Lake Bonneville. Journ. Geophys. Res., vol. 68, pp. 5517-5530.

DALY, R. A.

1934. The Changing World of the Ice Age. Yale University Press, New Haven, Conn., 271 pp.

DICKE, R. H.

1966. The secular acceleration of the earth's rotation and cosmology. Conference on the Earth-Moon System, NASA Inst. Space Studies, Plenum Press, New York, pp. 98-164.

DIMENT, W. H.

1954. Regional gravity survey in Vermont, western Massachusetts and eastern New York. Doctoral thesis, Harvard University.

EARDLEY, A. J.

1964. Polar rise and equatorial fall of sea level. Amer. Scientist, vol. 52, pp. 488-497.

FROLOV, A. I.

1964. The gravitational field and some features of crustal structure in Antarctica. Bull. Acad. Sci. USSR, Geophys. Ser. English Trans., no. 10, pp. 878-885.

GRIGGS, D. T., TURNER, F. J., AND HEARD, H. C.

1960. Deformation of rocks at 500° to 800° C. In Rock Deformation, Geol. Soc. Amer. Mem., vol. 79, pp. 39-104.

HASKELL, N. A.

1935. The motion of a viscous fluid under a surface load. 2. Physics, vol. 6, pp. 265-269.

HEINTZ, N.

1962. The Nansen ridge has become the Nansen trough. Norsk Polar-institutt Arbok, pp. 66-74.

HEISKANEN, W.

1936. Das Problem der Isostasie. Handbuch der Geophysik, 4th ed., vol. 1, Berlin.

HILL, R.

1950. The Mathematical Theory of Plasticity. Oxford Univ. Press, New York, 254 pp.

INNES, M. J. S.

1960. Gravity and isostasy in northern Ontario and Manitoba. Publ. Dominion Obs. Ottawa, vol. 21, pp. 265-338.

IZSAK, I. G.

1963a. Tesseral harmonics in the geopotential. Nature, vol. 199, pp. 137-139.

1963b. Tesseral harmonics in the geopotential. Private communication.

KORT, V. G.

1964. Antarctic oceanography. In Research in Geophysics, vol. 2, ed. by H. Odishaw, MIT Press, Cambridge, Mass., pp. 309-333.

KOZAI, Y.

1964. New determination of zonal harmonics coefficients of the earth's gravitational potential. Smithsonian Astrophys. Obs. Spec. Rep. No. 165, 38 pp.

KUENEN, Ph. H.

1950. Marine Geology. John Wiley and Sons, New York.

LEE, W. H. K.

1963. Heat flow data analysis. Rev. Geophys., vol. 1, pp. 449-479.

LEE, W. H. K., AND MAC DONALD, G. J. F.

1963. The global variation of terrestrial heat flow. Journ. Geophys. Res., vol. 68, pp. 6481-6492.

LOVEJOY, E. M. P.

1964. The problem of Neogene isostatic adjustments in Nevada (abstract), Trans. Amer. Geophys. Union, vol. 45, p. 638.

LOVERING, J. F., AND MORGAN, J. W.

1963. Uranium and thorium abundances in possible upper mantle material. Nature, vol. 197, pp. 138-140.

MUNK, W. H., AND MAC DONALD, G. J. F.

1960. Continentality and the gravitational field of the earth. Journ. Geophys. Res., vol. 65, pp. 2169-2172.

OSTENSO, N. A.

1962. Geophysical investigations of the Arctic Ocean basin. Wisc. Univ. Geophys. and Polar Res. Center Res. Rep. 62-4, 124 pp.

1963. Physiography and structure of the Arctic Ocean basin. Trans. Amer. Geophys. Union, vol. 44, pp. 637-648.

STRANGE, W. E.

1966a. Private communication.

1966b. A comparison of satellite gravity with surface gravity and other physical parameters (abstract). In Proceedings VIIth International COSPAR Space Sci. Symp., North-Holland Publ. Co., Amsterdam, in press.

TAKEUCHI, H.

1963. Time scales of isostatic compensations. Journ. Geophys. Res., vol. 68, p. 2357.

TAKEUCHI, H., AND HASEGAWA, Y.

1964. Viscosity distribution within the earth. Geophys. Journ., vol. 9, pp. 503-508.

THEIL, E. C.

1962. The amount of ice on planet earth. In Antarctic Research, Geophys. Mono. 7, ed. by H. Wexler, M. J. Rubin, and J. E. Coskey, Jr., Amer. Geophys. Union, pp. 172-175.

TILTON, G. R., AND REED, G. W.

1963. Radioactive heat production in eclogite and some ultramafic rocks. In Earth Science and Meteoritics, compiled by J. Geiss and E. D. Goldberg, North-Holland Publ. Co., Amsterdam, chapter 2.

VENING MEINESZ, F. A.

1934. Gravity expeditions at sea. Publ. Netherlands Geod. Comm., vol. 2, pp. 1-139.

WANG, C.-Y.

1963. On the distribution of surface heat flow and the second-order variations in the external gravitational field. Smithsonian Astrophys. Obs. Spec. Rep. No. 134, 13 pp.

WOOLLARD, G. P.

1962. Crustal structure in Antarctica. In Antarctic Research, Geophys. Mono. 7, ed. by H. Wexler, M. J. Rubin, and J. E. Coskey, Jr., Amer. Geophys. Union, pp. 53-73.

ZHIVAGO, A. V.

1962. Outlines of southern ocean geomorphology. In Antarctic Research, Geophys. Mono. 7, ed. by H. Wexler, M. J. Rubin, and J. E. Coskey, Jr., Amer. Geophys. Union, pp. 74-80.

DETERMINATION OF LOVE'S NUMBER FROM SATELLITE[1,2] OBSERVATIONS

Yoshihide Kozai[3]

The three satellites 1959 α1, 1959 η, and 1960 ι2 (see Table 1) were adopted for this study simply because their orbital elements determined from precisely reduced Baker-Nunn observations were available for long intervals of time. Of the orbital elements, the inclination of the orbital plane to the equator is the most accurately determined since, because of the geographical distribution of the SAO astrophysical observing stations, many observations are made near apices.

[1] Editors' Note: This paper was read at the Discussion Meeting on Orbital Analysis sponsored by the Royal Society, London, October 1966.

[2] Editors' Note: In addition to this topic, Dr. Kozai discussed in his seminar the possibility of detecting changes in J_2 caused by changes in the moments of inertia of the earth. In an article recently published, Kozai (1966) has the following remark:

> Besides the perturbations due to the tides, Kozai derived variations with one year period in the longitude of the ascending node for 1960 Iota 2. He thought that the variations were due to change in J_2 caused by change in the moments of inertia of the earth. And if this idea is correct, the amplitude of the variation of J_2 could be computed as 2×10^{-9}, from the data. However, more data are necessary before we can derive any definite conclusion.

[3] Astronomer, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts; regularly at Tokyo Astronomical Observatory, Mitaka, Tokyo, Japan.

Table 1. Satellites used in the determinations of Love's number

Satellite	a	i	e	$\dot{\Omega}$	Date
1959 α1 (Vanguard 2)	1.30	32.°8	0.16	-3.°5/day	Feb. 1959 - May 1964
1959 η (Vanguard 3)	1.33	33.4	0.19	-3.3	Sept. 1959 - Jan. 1964
1960 ι2 (rocket of Echo 1)	1.25	47.2	0.01	-3.1	Sept. 1960 - June 1964

The orbital elements are derived every 2 days by using observations of 4-day intervals. Every possible perturbation is computed and subtracted from the orbital elements.

Besides lunisolar, solar-radiation, and many other known perturbations, effects due to the motion of the earth's equator should be taken into account. That is, since the orbital elements are referred to the moving equator, which is not a reference plane of the inertial system, perturbations are produced in some of the orbital elements (Kozai, 1960).

The deformation of the earth due to the moon and the sun causes perturbations in satellite orbits (Kaula, 1962; Newton, 1965; Kozai, 1965). The most important perturbations for the inclinations of these three satellites are

$$0.°80 \times 10^{-3} k_2 \cos(\Omega - \nu \cdot \Delta t) \quad , \quad \text{for 1959 } \alpha 1 \text{ and 1959 } \eta$$

and

$$0.°79 \times 10^{-3} k_2 \cos(\Omega - \nu \cdot \Delta t) \quad , \quad \text{for 1960 } \iota 2 \quad ,$$

where k_2 is Love's number, Ω is the longitude of the ascending node of the satellite, ν is the angular rotational velocity of the earth, and Δt is the time lag of the tides due to the friction.

The perturbation terms with the argument Ω are produced by lunisolar gravitational attractions and the motion of the equator, and they should be computed carefully and subtracted from the inclination.

From the residuals of the observed inclinations every 2 days, the following periodic variations with argument Ω are derived:

$$\begin{matrix} 0.^{\circ}36 \times 10^{-3} \cos(\Omega - 11^{\circ}) \\ \pm\ 7 \qquad\qquad \pm\ 10 \end{matrix} \qquad \text{for 1959 } \alpha 1 \ ,$$

$$\begin{matrix} 0.^{\circ}38 \times 10^{-3} \cos(\Omega + 9^{\circ}) \\ \pm 11 \qquad\qquad \pm\ 12 \end{matrix} \qquad \text{for 1959 } \eta \ ,$$

$$\begin{matrix} 0.^{\circ}28 \times 10^{-3} \cos(\Omega + 23^{\circ}) \\ \pm\ 6 \qquad\qquad \pm\ 11 \end{matrix} \qquad \text{for 1960 } \iota 2 \ .$$

By comparing the observed amplitudes with the theoretical values, values for k_2 are computed as

$$k_2 = 0.45 \pm 0.09 \qquad \text{for 1959 } \alpha 1 \ ,$$

$$k_2 = 0.47 \pm 0.12 \qquad \text{for 1959 } \eta \ ,$$

$$k_2 = 0.35 \pm 0.06 \qquad \text{for 1960 } \iota 2 \ .$$

The mean value is

$$k_2 = 0.39 \pm 0.05 \ .$$

The time lag Δt for the tides can also be computed from the observed phase angle. However, since the scattering of Δt derived for the three satellites is large, Δt thus derived is not well determined, but is reasonably small.

REFERENCES

KAULA, W. M.

1962. Celestial geodesy. Adv. Geophys., vol. 9, pp. 191-293.

KOZAI, Y.

1960. Effect of precession and nutation on the orbital elements of a close earth satellite. Astron. Journ., vol. 65, pp. 621-623.

1965. Effect of the tidal deformation of the earth on the motion of close earth satellites. Publ. Astron. Soc. Japan, vol. 17, pp. 395-402.

1966. The earth's gravitational potential derived from satellite motion. Spac. Sci. Rev., vol. 5, pp. 818-879.

NEWTON, R. R.

1965. An observation of the satellite perturbation produced by the solar tide. Journ. Geophys. Res., vol. 70, pp. 5983-5989.

RELATIVISTIC INVESTIGATIONS

Brian G. Marsden[1] and James P. Wright[2]

1. INTRODUCTION

There is a statement attributed to Einstein, which closed a discussion in 1918 with H. Thirring, in which he lamented, "What a pity the earth has no moon in an orbit just outside its atmosphere" (with which to observe relativistic effects). There are now many "moons" just outside the earth's atmosphere, and it is important to discuss the magnitude of the relativistic effects. We shall not go deeply into general relativity theory, but we shall use the equations derived from the general theory.

Effects arise from force terms in the geodesic equations of motion that are not present in the Newtonian theory. For instance, there would be additional terms for a satellite in orbit about a rotating or nonrotating body; the field seen by a moving body is different from that seen by a body not moving relative to a central body; the motion of an earth satellite is slightly different because the earth is also in orbit around the sun.

The method of observation must also be considered, and reduced in a consistent manner. The observations usually do not involve distance, but only angle and time measurements. The observations by laser and radar are time-time measurements, and the distance is again inferred. The motion of a test particle or a light ray ($ds^2 = 0$) is determined by the geodesic equations that contain functions of the metric tensor g_{ij}.

The line element ds, or the "distance" between neighboring events in space and time, is given by

[1]Astronomer, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts.

[2]Astrophysicist, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts.

$$ds^2 = \sum_{j=1}^{4} \sum_{k=1}^{4} g_{jk}\, dx^j\, dx^k \quad ,$$

where three of the x^j denote spatial coordinates and the fourth the time (usually multiplied by the velocity of light for dimensional convenience). We have

$$g_{jk} = g_{kj} \quad ,$$

so that there are 10 independent g_{jk}, and in general they are functions of all the space and time coordinates.

A world line, which defines the motion of a particle in space and time, is obtained by finding the extremum of $\int ds$. We can thus obtain the equations of motion:

$$\frac{d^2 x^i}{ds^2} + \sum_{j=1}^{4} \sum_{k=1}^{4} \Gamma^i_{jk} \frac{dx^j}{ds} \frac{dx^k}{ds} = 0 \quad , \qquad (i = 1, 2, 3, 4)$$

where the Christoffel symbols Γ^i_{jk} are defined by

$$\Gamma^i_{jk} = \frac{1}{2} \sum_{p=1}^{4} g^{pi} \left(\frac{\partial g_{jp}}{\partial x^k} + \frac{\partial g_{kp}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^p} \right) \quad ;$$

the g^{pi} are the minors of the g_{pi}, divided by the determinant of the g_{pi}. A derivation of this is given in all the standard texts on relativity (e.g., Eddington, 1923, page 58; Landau and Lifshitz, 1962, § 98, § 113).

It is convenient to separate the time coordinate $x^4 = ct$ (c is the velocity of light) from the space coordinates x^1, x^2, x^3. Thus we obtain

$$\frac{d^2x^i}{dt^2} + \sum_{j=1}^{4}\sum_{k=1}^{4}\left[\Gamma^i_{jk} - \frac{1}{c}\,\Gamma^4_{jk}\,\frac{dx^i}{dt}\right]\frac{dx^j}{dt}\,\frac{dx^k}{dt} = 0 \quad . \quad (i = 1, 2, 3) \qquad (1)$$

If there were no mass in the universe, then we should have the flat-space metric

$$(g_{jk}) = (g^{jk}) = \begin{vmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \end{vmatrix} .$$

In general, it is convenient to take

$$(g_{jk}) = \begin{vmatrix} -1+\gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} \\ \gamma_{12} & -1+\gamma_{22} & \gamma_{23} & \gamma_{24} \\ \gamma_{13} & \gamma_{23} & -1+\gamma_{33} & \gamma_{34} \\ \gamma_{14} & \gamma_{24} & \gamma_{34} & +1+\gamma \end{vmatrix} ,$$

where the γ_{jk} are, we hope, small quantities. The most important effect arises from the γ (in the slow-motion approximation) in the expression for g_{44}, and the most important Christoffel symbols, therefore, are

$$\Gamma^i_{44} = -\frac{1}{2}\,g^{ii}\,\frac{\partial g_{44}}{\partial x^i} \quad . \qquad (i = 1, 2, 3)$$

Thus, our equations of motion (1) become

$$\frac{d^2x^i}{dt^2} = - c^2 \Gamma^i_{44} \quad ,$$

and since to the required order of approximation we have $g^{ii} = -1$ $(i \neq 4)$,

$$\frac{d^2x^i}{dt^2} = - \frac{1}{2} c^2 \frac{\partial \gamma}{\partial x^i} \quad . \tag{2}$$

Now we can compare these equations with the two-body problem, where we consider the motion of a point mass m with respect to a point mass M. If x^1, x^2, and x^3 are Cartesian coordinates of m with respect to M, then the equations of motion are

$$\frac{d^2x^i}{dt^2} = \frac{\partial U}{\partial x^i} \quad , \tag{3}$$

where the force function U is given by

$$U = \frac{\mu}{r} \quad ,$$

with

$$r^2 = \sum_{i=1}^{3} (x^i)^2$$

and

$$\mu = k^2 (M+m) \quad ,$$

k^2 being the constant of gravitation. Comparing the expressions (2) and (3), we find that

$$\gamma = -\frac{2\lambda}{r} \quad , \tag{4}$$

where

$$\lambda = \frac{\mu}{c^2}$$

and has the dimensions of length.

Generally, the g_{jk} have to be determined from the distribution of matter in space (and time). The solution is somewhat arbitrary, however, for although there are 10 g_{jk} to be determined, only 6 unique expressions exist for obtaining them. Thus, four arbitrary conditions appear in the definition of a metric.

It is often convenient to obtain three of these conditions from the assumption that the field around a mass is spherically symmetric and therefore static; hence, $g_{j4} = 0$ $(j \neq 4)$. Working in terms of polar coordinates, we can then write

$$ds^2 = -(1+\alpha)\, dr^2 - (1+\beta)\left(r^2 \sin^2\theta\, d\chi^2 + r^2\, d\theta^2\right)$$
$$+ (1+\gamma)\, c^2\, dt^2 \quad ,$$

where α, β, and γ are functions of r alone. We still have a choice for the remaining condition (and the coordinate conditions, usually expressed at infinity; see Fock (1959)), and that choice defines what we mean by a "length." The most commonly used metric is that due to Schwarzschild, and to the first power of λ we can write

$$\alpha = \frac{2\lambda}{r} \quad , \qquad \beta = 0 \quad , \qquad \gamma = -\frac{2\lambda}{r} \quad .$$

The value of γ corresponds with that in (4) to the second power of λ. Alternatively, the isotropic metric could be used, in which

$$\alpha = \frac{2\lambda}{r_1} \quad , \qquad \beta = \frac{2\lambda}{r_1} \quad , \quad \text{and} \quad \gamma = -\frac{2\lambda}{r_1} + \frac{2\lambda^2}{r_1^2} \quad .$$

(A coordinate transformation $r = r_1 \left[1 + (\lambda/2r_1)\right]^2$ transforms the Schwarzschild metric to the isotropic metric.) Converting the Schwarzschild metric to Cartesian coordinates $x^1 = x$, $x^2 = y$, and $x^3 = z$, we obtain

$$(g_{jk}) = \begin{vmatrix} -1 - \frac{2\lambda x^2}{r^3} & -\frac{\lambda xy}{r^3} & -\frac{\lambda zx}{r^3} & 0 \\ -\frac{\lambda xy}{r^3} & -1 - \frac{2\lambda y^2}{r^3} & -\frac{\lambda yz}{r^3} & 0 \\ -\frac{\lambda zx}{r^3} & -\frac{\lambda yz}{r^3} & -1 - \frac{2\lambda z^2}{r^3} & 0 \\ 0 & 0 & 0 & 1 - \frac{2\lambda}{r} \end{vmatrix} \quad .$$

2. RELATIVISTIC EQUATIONS FOR THE TWO-BODY PROBLEM SEEN AS A PERTURBATION TO THE NEWTONIAN EQUATIONS

After working out the Christoffel symbols and substituting them into (1), we obtain, instead of (3), the equations of motion

$$\frac{d^2x^i}{dt^2} = \frac{\partial U}{\partial x^i} + X^i \quad , \qquad (i = 1, 2, 3) \qquad (5)$$

where

$$X^i = Ax^i + B\dot{x}^i \quad ; \qquad (6)$$

$\dot{x}^i$ denotes dx^i/dt,

$$A = \frac{\lambda}{r^3}\left(\frac{2\mu}{r} - 2v^2 + 3\dot{r}^2\right) \quad \text{and} \quad B = \frac{2\lambda\dot{r}}{r^2} \quad , \qquad (7)$$

where

$$v^2 = \sum_{i=1}^{3} (\dot{x}^i)^2 \quad \text{and} \quad \dot{r} = \frac{1}{r}\sum_{i=1}^{3} x^i\dot{x}^i \quad .$$

These equations can be compared with those given by Brouwer and Clemence (1961a, p. 49). If we now take

$$F = U - \frac{1}{2} v^2 \quad ,$$

then (5) can be written

$$\frac{d\dot{x}^i}{dt} = \frac{\partial F}{\partial x^i} + X^i \quad ; \quad \frac{dx^i}{dt} = - \frac{\partial F}{\partial \dot{x}^i} \quad . \qquad (i = 1, 2, 3) \qquad (8)$$

In the nonrelativistic case, the X^i are zero and these equations are canonical. The Hamiltonian $-F$ is the total energy of the system, equal to $-\mu/2a$, where a is the semimajor axis of the orbit. We may transform from the canonical set $(\dot{x}^i, x^i)$ to the set (L_i, ℓ_i), given by

$$L_1 = \sqrt{\mu a} \qquad \ell_1 = \text{mean anomaly},$$

$$L_2 = L_1 \sqrt{1 - e^2} \qquad \ell_2 = \text{argument of the pericenter},$$

$$L_3 = L_2 \cos I \qquad \ell_3 = \text{longitude of the ascending node},$$

where e and I are the orbital eccentricity and the inclination of the orbit to the fundamental plane. These are the well-known Delaunay variables: In the nonrelativistic two-body problem, all are constant except for ℓ_1, which is a linear function of time. In the nonrelativistic n-body problem, we can still obtain a Hamiltonian $-F$, but the final expressions for L_i and ℓ_i are more complex.

The additional complication of the X^i can be handled quite simply (Brouwer and Clemence, 1961b, p. 577). Multiplying the first of (8) by $\partial x^i / \partial \ell_j$, the second by $- \partial \dot{x}^i / \partial \ell_j$, and adding, we get

$$\frac{d\dot{x}^i}{dt} \frac{\partial x^i}{\partial \ell_j} - \frac{dx^i}{dt} \frac{\partial \dot{x}^i}{\partial \ell_j} = \frac{\partial F}{\partial x^i} \frac{\partial x^i}{\partial \ell_j} + \frac{\partial F}{\partial \dot{x}^i} \frac{\partial \dot{x}^i}{\partial \ell_j} + X^i \frac{\partial x^i}{\partial \ell_j} \quad . \tag{9}$$

Now we have

$$\frac{d\dot{x}^i}{dt} = \sum_{k=1}^{3} \left(\frac{\partial \dot{x}^i}{\partial L_k} \frac{dL_k}{dt} + \frac{\partial \dot{x}^i}{\partial \ell_k} \frac{d\ell_k}{dt} \right)$$

and

$$\frac{dx^i}{dt} = \sum_{k=1}^{3} \left(\frac{\partial x^i}{\partial L_k} \frac{dL_k}{dt} + \frac{\partial x^i}{\partial \ell_k} \frac{\partial \ell_k}{dt} \right) \quad .$$

Substituting into (9) and summing over i, we get

$$\sum_{k=1}^{3} \left\{ [\ell_j, L_k] \frac{dL_k}{dt} + [\ell_j, \ell_k] \frac{d\ell_k}{dt} \right\} = \frac{\partial F}{\partial \ell_j} + \sum_{i=1}^{3} X^i \frac{\partial x^i}{\partial \ell_j} \quad , \tag{10}$$

where the notation of Lagrange brackets is used, namely,

$$[p, q] = \sum_{i=1}^{3} \left(\frac{\partial x^i}{\partial p} \frac{\partial \dot{x}^i}{\partial q} - \frac{\partial x^i}{\partial q} \frac{\partial \dot{x}^i}{\partial p} \right) \quad .$$

Similarly, we obtain

$$\sum_{k=1}^{3} \left\{ [L_j, L_k] \frac{dL_k}{dt} + [L_j, \ell_k] \frac{d\ell_k}{dt} \right\} = \frac{\partial F}{\partial L_j} + \sum_{i=1}^{3} X^i \frac{\partial x^i}{\partial L_j} \quad . \tag{11}$$

Since the Delaunay variables are canonical, all the Lagrange brackets are zero, except for

$$[\ell_j, L_j] = 1 \qquad \text{and} \qquad [L_j, \ell_j] = -1 \quad .$$

Thus, (10) and (11) become

$$\frac{dL_j}{dt} = \frac{\partial F}{\partial \ell_j} + P_j \qquad \text{and} \qquad \frac{d\ell_j}{dt} = -\frac{\partial F}{\partial L_j} - Q_j \quad , \tag{12}$$

where

$$P_j = \sum_{i=1}^{3} X^i \frac{\partial x^i}{\partial \ell_j} \qquad \text{and} \qquad Q_j = \sum_{i=1}^{3} X^i \frac{\partial x^i}{\partial L_j} \quad . \tag{13}$$

By analogy with (6), we write

$$\begin{aligned} P_j &= Ap_j + Bp'_j \quad , \\ Q_j &= Aq_j + Bq'_j \quad , \end{aligned} \tag{14}$$

where

$$p_j = \sum_{i=1}^{3} x^i \frac{\partial x^i}{\partial \ell_j} \qquad p'_j = \sum_{i=1}^{3} \dot{x}^i \frac{\partial x^i}{\partial \ell_j}$$

$$q_j = \sum_{i=1}^{3} x^i \frac{\partial x^i}{\partial L_j} \qquad q'_j = \sum_{i=1}^{3} \dot{x}^i \frac{\partial x^i}{\partial L_j} \quad .$$

Specifically, we have

$$p_1 = \frac{L_1}{n} e \sin u \qquad p'_1 = L_1 \frac{1 + e \cos u}{1 - e \cos u}$$

$$p_2 = 0 \qquad p'_2 = L_2$$

$$p_3 = 0 \qquad p'_3 = L_3$$

$$q_1 = \frac{1}{ne}\left[e(3-e^2) - (1+3e^2)\cos u + 2e^3 \cos^2 u\right] \qquad q'_1 = \frac{2(1-e^3\cos u)}{e} \frac{\sin u}{1-e\cos u}$$

$$q_2 = \frac{\sqrt{1-e^2}}{ne}(\cos u - e) \qquad q'_2 = -\frac{2\sqrt{1-e^2}}{e} \frac{\sin u}{1-e\cos u}$$

$$q_3 = 0 \qquad q'_3 = 0 \quad , \qquad (15)$$

where n is the mean motion and u the eccentric anomaly. When the eccentricity and the inclination are small, the expressions (7) and (15) can be expanded in powers of these quantities. Hence, determining P_j and Q_j from (14), substituting into (12), and integrating, we obtain for the perturbations in the Delaunay variables, correct to the second order:

$$\frac{\delta L_1}{eL_1} = \frac{\lambda}{a}\left[\left(-2-e^2\right)\cos \ell_1 - \frac{5}{2} e \cos 2\ell_1 - \frac{7}{3} e^2 \cos 3\ell_1\right]$$

$$\frac{\delta L_2}{eL_1} = \frac{\lambda}{a}\left[\left(-2+\frac{5}{4} e^2\right)\cos \ell_1 - 2e \cos 2\ell_1 - \frac{9}{4} e^2 \cos 3\ell_1\right]$$

$$\frac{\delta L_3}{eL_1} = \frac{\lambda}{a}\left[\left(-2+\frac{5}{4} e^2 + I^2\right)\cos \ell_1 - 2e \cos 2\ell_1 - \frac{9}{4} e^2 \cos 3\ell_1\right]$$

$$\delta \ell_1 = \frac{\lambda}{a}\left[\left(-3-\frac{9}{2} e^2\right)nt - \frac{17}{4} e \sin \ell_1 + \left(\frac{1}{2} - \frac{1}{2} e^2\right)\sin 2\ell_1 + \frac{3}{4} e \sin 3\ell_1 + \frac{7}{8} e^2 \sin 4\ell_1\right]$$

$$\delta \ell_2 = \frac{\lambda}{a}\left[(3+3e^2)\, nt + \frac{33}{4} e \sin \ell_1 + \left(-\frac{1}{2} + \frac{23}{4} e^2\right)\sin 2\ell_1 - \frac{3}{4} e \sin 3\ell_1 - \frac{7}{8} e^2 \sin 4\ell_1\right]$$

$$\delta \ell_3 = 0 \quad .$$

It is perhaps useful to convert these to perturbations in the Keplerian elements a, e, and I, the mean longitude $\hat{\lambda}$, and the longitude of pericenter $\bar{\omega}$; we also use the notation Ω for the longitude of the ascending node and ℓ for the mean anomaly. It will be found that

$$\frac{\delta a}{a} = \frac{\lambda}{a}\left(-4e \cos \ell - 5e^2 \cos 2\ell\right)$$

$$\delta e = \frac{\lambda}{a}\left(-\frac{5}{4} e^2 \cos \ell - \frac{1}{2} e \cos 2\ell - e^2 \cos 3\ell\right)$$

$$\delta I = 0$$

$$\delta\hat{\lambda} = \frac{\lambda}{a}\left(-\frac{3}{2}e^2 nt + 4e \sin \ell + \frac{21}{4} e^2 \sin 2\ell\right)$$

$$\delta\bar{\omega} = \frac{\lambda}{a}\left[\, 3(1 + e^2)\, nt + \frac{33}{4} e \sin \ell - \frac{1}{4}(2 - 23e^2) \sin 2\ell \right.$$
$$\left. - \frac{3}{4} e^2 \sin 3\ell - \frac{7}{8} e^2 \sin 4\ell \right]$$

$$\delta\Omega = 0 \quad . \tag{16}$$

The pericenter is found to have its well-known secular perturbation $(3\lambda/p)\,nt$, where p is the semilatus rectum. The more standard derivation of this perturbation is given, for example, by Danby (1962, p. 66). It also appears that there is a secular term in the mean longitude. This term is actually not there, however, because the secular part of $d\hat{\lambda}/dt$ is the mean motion, by definition. For the mean motion we must take

$$\bar{n} = n\left(1 - \frac{3}{2} e^2 \frac{\lambda}{a}\right) \quad ,$$

where n is the "mean motion" consistent with the semimajor axis by Kepler's third law:

$$n^2 a^3 = \mu \quad .$$

It remains to calculate the effects of relativity on the true longitude ψ and the radius vector r. From (16) it follows that

$$\delta\psi = \frac{\lambda}{a}\left(-6e\, nt \cos \ell - \frac{15}{2} e^2 nt \cos 2\ell + 5e \sin \ell \right.$$
$$\left. + \frac{3}{2} e^2 \sin 2\ell - \frac{1}{4} e^2 \sin 4\ell\right)$$

$$\frac{\delta r}{a} = \frac{\lambda}{a}\left(-3ent\sin\ell - 3e^2 nt\sin 2\ell + e^2 - \frac{7}{2}e\cos\ell + \frac{1}{8}e^2\cos 2\ell + \frac{1}{8}e^2\cos 4\ell\right) \quad . \tag{17}$$

There appears to be a constant perturbation $e^2(\lambda/a)$ in the radius vector. This is inseparable from a constant perturbation in the semimajor axis and is unobservable. We must define instead the mean distance

$$\bar{a} = a\left(1 + e^2\frac{\lambda}{a}\right) \quad .$$

In passing, we note that Kepler's third law holds if $\bar{n}$ and $\bar{a}$ are used instead of n and a: certainly the terms in the first power of $e^2(\lambda/a)$ disappear. This is not the case if the isotropic metric is used, but any departure from Kepler's third law would be undetectable. The effect on a and n is due to the choice of coordinates, as the distance "r" is a variable used to relate angle and time.

In the nonrelativistic two-body problem the coefficients of sin ℓ in the true longitude and cos ℓ in the radius vector are in the ratio 2:-1. In the expression (17) for the perturbations, the corresponding coefficients are almost in the same ratio. If the terms were instead $+6e(\lambda/a)\sin\ell$ and $-3e(\lambda/a)\cos\ell$, they would be in the same ratio and undetectable as perturbations, for the same effect would be obtained if, rather than e, we used the mean eccentricity

$$\bar{e} = e\left(1 + 3\frac{\lambda}{a}\right) \quad .$$

Thus, when discussing effects of relativity that might be detectable, we must take instead of (17) the expressions

$$\delta\psi = \frac{\lambda}{a}\left(-6ent\cos\ell - \frac{15}{2}e^2 nt\cos 2\ell - e\sin\ell + \frac{3}{2}e^2\sin 2\ell - \frac{1}{4}e^2\sin 4\ell\right)$$

$$\frac{\delta r}{a} = \frac{\lambda}{a}\left(-3ent\sin\ell - 3e^2 nt\sin 2\ell - \frac{1}{2}e\cos\ell + \frac{1}{8}e^2\cos 2\ell + \frac{1}{8}e^2\cos 4\ell\right) .$$

3. POSSIBILITY OF OBSERVING RELATIVISTIC EFFECTS

When the primary body is taken to be the sun, the length λ is 10^{-8} a.u., or 1.5 km. For the orbit of Mercury, the secular motion $\delta\bar{\omega}$ is 43 arcsec century^{-1} (see Table 1) (Ginzburg, 1959, 1962).

When the primary body is taken to be the earth, λ is (1.5/333000) km = 4.5 mm. For the orbit of the moon, the secular motion $\delta\bar{\omega}$ is 0.06 arcsec century^{-1} (see Table 2). For an artificial satellite that just clears the earth's surface at perigee (ignoring the atmosphere), we might take $e = 0.2$, $a = 1.25$ earth radii, and $n = 77$ rad day^{-1}. Then the secular motion $\delta\bar{\omega}$ is 0.028 arcsec day^{-1}. The principal contributions of this secular term to the true longitude and radius vector are consequently the mixed secular terms

$$\delta\psi = -0.011\, t\cos\ell \text{ arcsec}$$

and

$$\delta r = -21\, t\sin\ell \text{ cm} ,$$

where t is measured in days. It would seem that the relativistic effect should be detectable, particularly in the radius vector, if use is made of lasers or radar. The figure of the earth also produces a secular motion, on the order of a few degrees per day, in the perigee of the orbit of a satellite. It should be possible to separate the small relativistic effect from it, however, because the figure also produces a secular motion in the node. If a value of the earth's dynamical form factor J_2 is to be determined from observations of an artificial satellite, it should be obtained from the motion of the node, or from the motion of the perigee after the latter has been corrected for the effect of relativity. For a satellite in an equatorial orbit, the J_2 component of the earth's field gives an advance of perigee

$$\Delta\bar{\omega} = \frac{3}{2} J_2 \left(\frac{R}{a}\right)^2 (1 - e^2)^{-2} nt \quad ,$$

where R is the earth's equatorial radius. A determination of J_2 is given as $(1082645 \pm 6) \times 10^{-9}$ (Kozai, 1964). For the satellite in the example, the relativistic correction would be equivalent to a change of 1.7×10^{-9} in J_2 in calculating the perigee. This is almost one-third the quoted error, but the correction would be three times the error if the observational accuracy were improved by an order of magnitude.

Table 1. Centennial advance of perihelion

	$\delta\bar{\omega}$	$e\delta\bar{\omega}$	$\delta\bar{\omega}$ observed
Mercury	43".03	8".847	42".56 ± 0".94
Venus	8.63	0.059	
Earth	3.84	0.064	4.6 ± 2.7
Mars	1.35	0.126	

Table 2. Centennial advance of perigee

	Distance (miles)	e	$\delta\bar{\omega}$	$e\delta\bar{\omega}$
Moon	240,000	0.06	0!!06	0!!0036
I	4200	0.01	1450	15
II	10,500	0.06	146	9
III	10,500	0.40	195	78
IV	4500	0.02	1250	25
V	6200	0.25	587	147

The principal periodic terms in ψ and r due to relativity also have the period of the mean anomaly, but their coefficients do not increase with time. In the example above,

$$\delta\psi = - 0.000023 \sin \ell \text{ arcsec}$$

and

$$\delta r = -0.045 \cos \ell \text{ cm} \quad ,$$

which are well below the limit of detectability.

If we use laser or radar methods to determine the "distance," we must also use the relativistic equation of motion of a light ray $ds^2 = 0$, plus the geodesic equation. If we use a radial light ray ($d\chi = 0$, $d\theta = 0$) and the Schwarzschild metric, then

$$ds^2 = 0 = \left(\frac{1 - 2\lambda}{r}\right) c^2 dt^2 - \left(\frac{1 - 2\lambda}{r}\right)^{-1} dr^2 \quad ,$$

or

$$c\, dt = \left(\frac{1 - 2\lambda}{r}\right)^{-1} dr \quad .$$

Integrating, we find

$$c\,\delta t \simeq r_2 - r_1 + 2\lambda \ln \frac{r_2}{r_1} \quad .$$

For an earth satellite with $r_2 = 10\, r_1$, $c\,\delta t - (r_2 - r_1) = 2.1$ cm; the correction would be less for closer orbits.

4. COMBINED EFFECT OF THE EARTH AND THE SUN ON THE MOTION OF A SATELLITE

The complete solution for the relativistic effects of the combined fields of the earth and the sun on a satellite of the earth is very complex. The metric must involve both the principal masses, and since they are revolving around each other, the field is no longer static, so that the g_{j4} terms now differ from zero. It was shown by de Sitter (1916a, b) that the motion of a satellite could be derived with a great deal of precision by considering the effects of the earth and the sun separately, for the cross terms give a negligible contribution.

We have already considered the part due to the earth. For that part due to the sun we must take into account the differential effects on the earth and the satellite. The principal contribution is given by (5), where U now includes terms due to the nonrelativistic attraction of the sun, and

$$X^i = 2\lambda\mathcal{M}\left[\mu\mathcal{M}\left(\frac{x^{i''}}{r''^4} - \frac{x^{i'}}{r'^4}\right) + \left(\frac{\dot{r}''\dot{x}^{i''}}{r''^2} - \frac{\dot{r}'\dot{x}^{i'}}{r'^2}\right) - \left(\frac{x^{i''}\,v''^2}{r''^3} - \frac{x^{i'} - v'^2}{r'^3}\right)\right] \quad .$$

Here, the single primes denote the heliocentric coordinates of the earth; the double primes, the heliocentric coordinates of the satellite; and $\mathcal{M}$, the sun-earth mass ratio.

If we take the orbit of the earth to be circular and the orbit of the satellite to be inclined at only a small angle to the ecliptic, the chief contribution to the secular terms is given by

$$X^3 = 3\frac{\lambda}{a'}\mathcal{M} n' n x^3 \quad ,$$

where n′ and a′ denote the mean motion and the mean distance of the earth's orbit about the sun. Instead of (13), we have merely

$$P_j = X^3 \frac{\partial x^3}{\partial \ell_j} \qquad \text{and} \qquad Q_j = X^3 \frac{\partial x^3}{\partial L_j} \quad .$$

As before, the P_j are sine functions and do not produce any secular terms upon integration. All we are left with, then, is

$$\frac{d\ell_3}{dt} = -Q_3 = \frac{3}{2}\frac{\lambda}{a'}\mathcal{M} n' \quad .$$

Hence,

$$\delta\bar{\omega} = \delta\Omega = \frac{3}{2}\frac{\lambda}{a'}\mathcal{M} n' t \quad .$$

There is also an effect in $\delta\hat{\lambda}$, but that cannot be separated from the mean motion. Numerically, the effect amounts to 1.9 arcsec century^{-1}. It appears in the motion of each artificial satellite and that of the moon, but it is not detectable by present techniques. We mention it here because there has been some confusion of this result with the phenomenon of geodesic precession (Eckert, 1965).

5. EFFECT OF THE ROTATION OF THE EARTH

The field about a spinning body is different from that of a nonrotating body, and in general relativity this also produces an advance of the pericenter. The ratio of the Lense-Thirring effect to the ordinary relativistic advance is (assuming a solid homogeneous body)

$$\frac{\delta\bar{\omega}_{LT}}{\delta\bar{\omega}_{GR}} = \frac{8}{15}\left(\frac{r_0}{a}\right)^2\left(\frac{T}{\tau}\right) \quad ,$$

where τ is the period of the spinning body, T the period of the orbit, r_0 the radius of the body, and a the radius of the orbit. For a close earth satellite, the ratio would be 3.3×10^{-2} and may be neglected.

6. CONCLUSION

The magnitudes of the relativistic corrections to satellite orbits are now just smaller than the proposed 1-meter accuracy envisaged in the immediate future. However, it will be important to separate the relativistic effects from those due to other causes, such as the nonsphericity of the earth, the influence of high-altitude winds, magnetic fields, and additional problems stemming from imprecise geodetic knowledge.

7. REFERENCES

BROUWER, D. and CLEMENCE, G. M.

1961a. Orbits and masses of planets and satellites. In The Solar System, Planets and Satellites, vol. 3, ed. by G. P. Kuiper and B. M. Middlehurst, Univ. Chicago Press, Chicago, pp. 31-94.

1961b. Methods of Celestial Mechanics. Academic Press, New York, 598 pp.

DANBY, J. M. A.

1962. Fundamentals of Celestial Mechanics. Academic Press, New York, 348 pp.

ECKERT, W. J.

1965. On the motions of the perigee and node and the distribution of mass in the moon. Astron. Journ., vol. 70, pp. 787-792.

EDDINGTON, A. S.

1923. The Mathematical Theory of Relativity. Cambridge Univ. Press, Cambridge, England, 270 pp.

FOCK, V.

1959. The Theory of Space, Time, and Gravitation. Pergamon Press, New York, 411 pp.

GINZBURG, V. L.

1959. Artificial satellites and the theory of relativity. Scientific Amer., vol. 200, pp.149-160.

1962. Experimental Verification of the General Theory of Relativity. In Recent Developments in General Relativity (Infeld Commemorative Volume), Pergamon Press, London, pp. 57-71.

KOZAI, Y.

1964. New Determination of zonal harmonics coefficients of the earth's gravitational potential. Publ. Astron. Soc. Japan, vol. 16, pp. 263-284.

LANDAU, L. D., and LIFSHITZ, E. M.

1962. The Classical Theory of Fields, 2nd ed., Addison-Wesley Publ. Co., Reading, Mass., 404 pp.

SITTER, W. DE

1916a. On Einstein's theory of gravitation and its astronomical consequences. Monthly Notices Roy. Astron. Soc., vol. 76, pp. 699-728.

1916b. On Einstein's theory of gravitation and its astronomical consequences. Monthly Notices Roy. Astron. Soc., vol. 77, pp. 155-184.

THE MOTION OF THE SPIN AXIS AND THE ROTATION OF THE EARTH

George Veis[1]

1. INTRODUCTION

A reference system is vital for studies in astronomy, geodesy, and geophysics, whether they are of geometric or of dynamic nature. For geometric studies, the reference system can be more or less arbitrarily selected, but for dynamic studies, an inertial reference system is advantageous. For mechanics related to the space close to the earth, a system defined by the center of mass (CM) of the earth and the equatorial plane will be satisfactory; but such a definition of a reference system is only theoretical, and in practice the problem is quite different. A reference system is defined in reality by a set of nominal coordinates given to some physical points used as reference. Here, of course, we assume that these coordinates are internally consistent, i. e., they satisfy the necessary metric properties.

Accordingly, for earth points, the reference system is actually defined by a set of "nominal" coordinates given to a number of points on the earth, while for space directions, the reference frame is defined by a set of nominal coordinates given to a number of stars. The frame for the terrestrial system is actually defined by the nominal astronomic latitudes of the International Polar Motion Service (IPMS) station and the nominal astronomic longitudes of the Bureau International de l'Heure (BIH) stations. For the celestial system, it has been agreed to use the coordinates of the FK-4 stars. The SAO Star Catalog is in the FK-4 system.

[1] Geodesist, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts; regularly at National Technical University, Athens, Greece.

A fundamental problem, therefore, is to find the relative position of those two basic systems. However, this relative position is not permanent, but changes with time. This change consists primarily of a systematic part that, to a major extent, can be predicted from theory. There remains, though, a small part that does not fit our present models, and that we try to explain with new theories. There is also a change that is not, so far as the reference system is concerned, of a systematic nature. It arises from proper motions of each of the individual points defining the system. These proper motions, which also consist of a systematic and a random part, are quite small; they are of the order of centimeters per year for the stations and of 0.01 arcsec year^{-1} for the stars (i.e., of the order of 10^{-7} or 10^{-8}). However, they introduce a rather important problem in the analysis, since these motions must be filtered out.

2. THE MOTION OF THE SPIN AXIS IN INERTIAL SPACE

A basic direction that plays an important role in problems related to the reference systems is the direction of the spin axis of the earth. This direction is important from an observational point of view, because it is easily observable. A camera photographing the stars as they rotate around the pole can easily define the direction of the pole. From a theoretic-dynamic point of view, it is also significant because it gives the direction of the ω vector.

Another important direction is the local vertical, or the direction of free fall. This direction is not fixed either in space or with respect to the crust of the earth, since it is affected by the direct and indirect tidal effects. The amplitude of these effects is approximately 0.02 arcsec (or 10^{-7}), which is not always negligible. The great advantage of using this direction is that it can be easily and precisely materialized (e.g., with a mercury bath).

The astronomic coordinates of a point are defined with these two directions, the spin axis and the local vertical. With modern technology, astronomic coordinates at an observatory can be determined over one night's observations to an accuracy better than ± 0.05 arcsec.

The celestial reference system, which is, realistically speaking, defined by the background stars, provides an inertial system for directions. In this system, the spin axis of the earth is not fixed. During the 2nd century B.C., Hipparchus discovered a slow rotational motion of the pole around the pole of the ecliptic. The period of this rotation was found to be approximately 26,000 years. This motion, known as precession, is due to the lunisolar attraction on the earth's equatorial bulge. There is an additional motion of the pole of the ecliptic, which rotates around the pole of the equator and which is due to the gravitational action of the planets on the earth. This

motion is known as planetary precession. The combined effect is known as general precession. As a result of precession, the positions of the ecliptic and of the equator – and thus of their intersection, the point γ – are continuously changing on the fixed-star background, i. e., in the celestial system.

Bradley, in the 18th century, discovered that the motion of the equatorial and the ecliptic planes had, in addition to the secular variation, a periodic variation as well, called nutation. The origin of nutation is the same as precession. It arises from the periodic variations in the gravitational attraction of the earth from the sun and the moon.

Modern theory provides over 60 periodic terms with a variety of amplitudes down to 0. 0001 arcsec. The most important terms are

A. one with an 18-year period with an amplitude of 13 arcsec, arising from the motion of the node of the moon's orbit,

B. one of 1/2 year with an amplitude of 1. 3 arcsec, arising from the sun's longitude,

C. one of 1/2 month with an amplitude of 0. 2 arcsec, arising from the longitude of the moon.

As a result of precession and nutation, the motion of the earth's spin axis in the inertial celestial reference system is quite complicated. Figures 1 and 2 give, at two different scales, the main features of this motion.

If we consider only the secular part of the motion of the ecliptic and the equator, i. e., the effect of precession alone, we introduce a fictitious mean equator and ecliptic in contrast to the true equator and ecliptic, which are obtained by adding also the periodic effect of nutation. The positions of stars referred to the mean equator and ecliptic of a certain date are called mean places, while positions referred to the true equator and ecliptic are called true places. The celestial system is defined with the mean places of stars at a certain epoch; e. g., the FK-4 system is defined for 1950. 0 epoch. Since

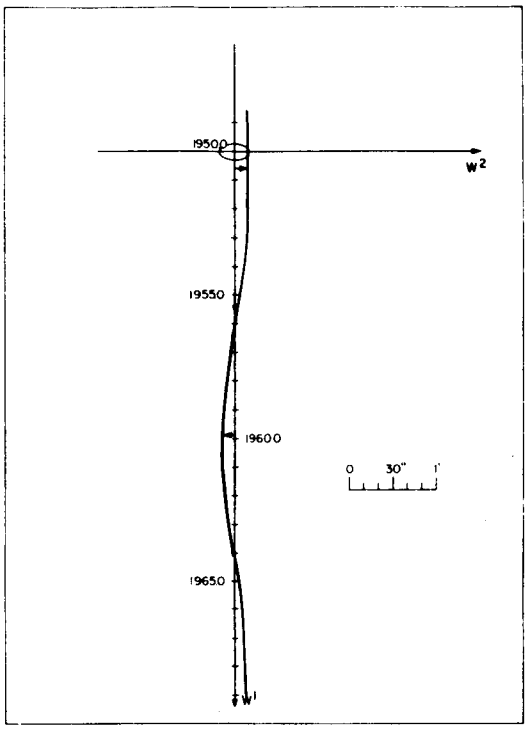


Figure 1. Precessional and nutational motion.

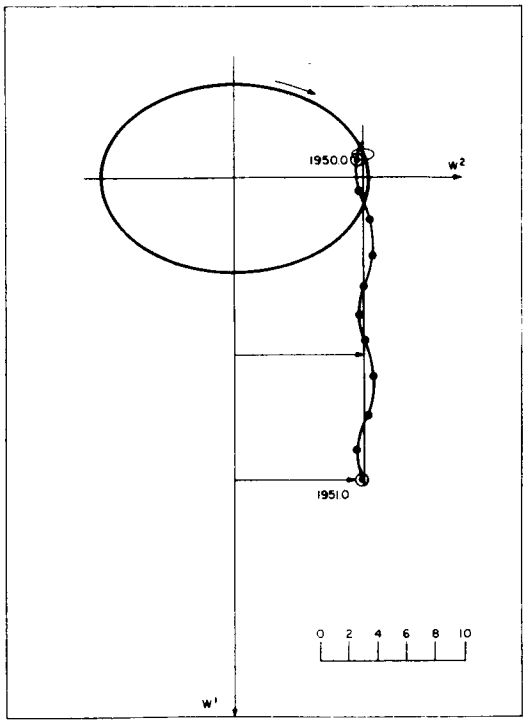


Figure 2. Precessional and nutational motion.

the earth's true axis of rotation is related to the true position of stars, it is apparent that the transformation from mean places to true places gives the actual orientation of the earth's spin axis in the inertial celestial system.

The transformation from the mean positions of a certain epoch to the true position of the date is given by the dynamic theory of precession and nutation. This theory depends, however, on some constants involved with the ratio (C - A)/C of the earth, the masses and the parallaxes of the sun and the moon, etc. There are also the constants of integration. All the constants must be determined from observations. There are only three independent constants: the obliquity at a certain epoch, the general precession in longitude, and the coefficient of the cos Ω term in the nutation in obliquity, which is the semimajor axis of the nutation eclipse (Figures 1 and 2).

3. THE MOTION OF THE SPIN AXIS IN THE EARTH'S BODY

Euler had already suggested, in 1765, that if the spin axis and the inertial axis of the earth did not coincide, there should be a free nutation that would, as a result, move the rotation axis of the earth continuously within its body. Euler predicted a rotation of the spin axis around a mean position with a period of 10 months (305 days).

In 1884, Künster found a periodic variation in the astronomic latitude of Potsdam, and in 1891, joint observations at Potsdam and Waĭkiki (which are in antidiametrical meridians) proved that the polar axis was indeed moving with a period of a little over a year. Owing to the importance of this phenomenon, the International Association of Geodesy established, in 1895, the International Latitude Service (ILS) consisting of five observatories all at latitude 38° 08' but at various longitudes. The purpose of ILS was to monitor continuously the position of the true terrestrial pole by determining the astronomic latitude every night at the participating observatories.

Let ϕ_0, λ_0 be the mean (or nominal) astronomic latitude and longitude of a station (which refer to the mean pole P_0), P the true pole, and ϕ, λ the true astronomic latitude and longitude (see Figure 3). If x, y are the coordinates of the true pole, we have

$$\phi = \phi_0 + x \cos \lambda_0 - y \sin \lambda_0 \tag{1}$$

and

$$\lambda = \lambda_0 + (x \sin \lambda_0 + y \cos \lambda_0) \tan \phi_0 \quad . \tag{2}$$

The determination of the astronomic latitude at a station for which the mean latitude is assumed to be known will give one equation of the form (1), relating the coordinates x and y of the true pole. Observations from two stations will give two equations that, provided the longitudes of the stations

are different, will give a unique solution for x and y. It must be emphasized that in this case the pole of reference used as the nominal mean pole is determined from the nominal mean latitudes of the two stations. Assuming that the two stations do not move, this method can monitor the motion of the pole quite well. If more than two stations are involved, the values of x and y will be derived from a least-squares solution. This method of determining the position of the pole is nothing else than an inversion of the well-known method of the Sumner line used in astronomic navigation.

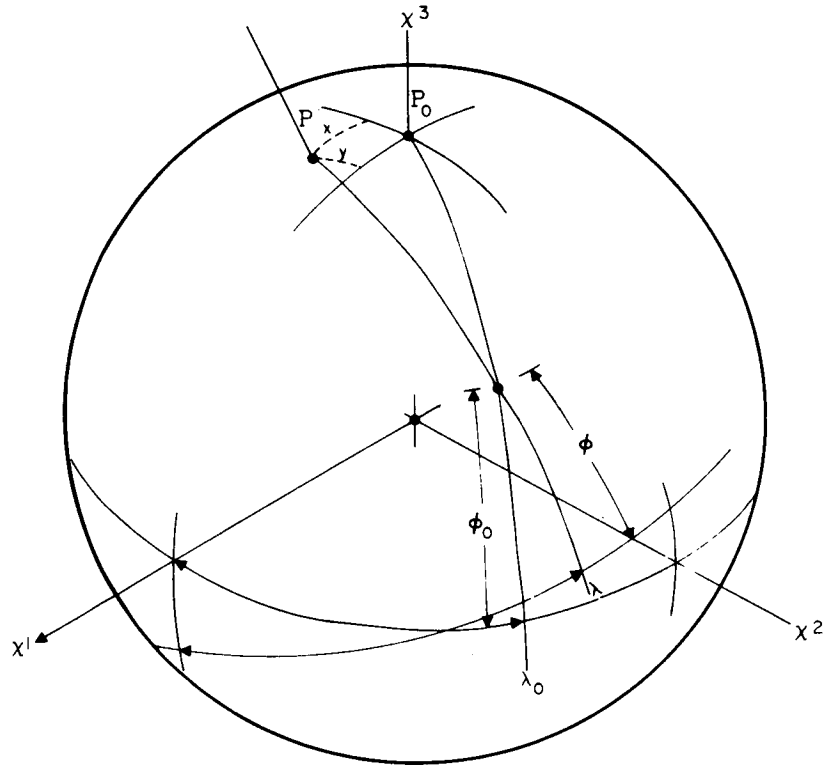


Figure 3. The effect of polar motion on astronomic coordinates.

To eliminate systematic errors, such as those arising from errors in the proper motion and the declinations of the stars used, from errors in the constants of aberration and nutation, etc., Kimura has introduced one more unknown z in equation (1), very often called the Kimura term or the z term. The introduction of the z term — which is of the order of 0.03 arcsec — corresponds, in replacing the intersection of the Sumner lines, with an

inscribed circle of radius z. There is some indication suggesting that the z term is the same in both hemispheres, but this is not really established as yet.

The ILS has continuous records for the polar motion since the service was established more than 60 years ago. Figure 4 gives this motion in recent years. Unfortunately, there have been several changes in the directorship of the ILS, which have also introduced changes in the observing programs, the catalogs of reference stars, the method of reduction, etc. As a result, it becomes quite a problem to reduce all available data to a homogeneous system.

Since 1960, the ILS has been replaced by the International Polar Motion Service (IPMS) with headquarters in Mizusawa, Japan. Currently, the determination of the polar motion is based not only on observations from the five old ILS stations, but also on those from several other observatories where a variety of instruments and methods are used. The variety of instruments, of course, creates even more problems for the homogeneous reduction of the data.

In spite of the inhomogeneity of the data, several conclusions can be drawn. In 1891, Chandler, a merchant from Cambridge, Mass., found that the polar motion is affected by two important periodic terms. Their periods are of 12 and 14 months. Figure 5 shows the spectrum of the polar motion (based on recent data), from which the two periods can easily be seen. The 14-month period is fairly flat, but the 12-month one has a rather sharp peak. The amplitudes of these two periodic terms are changing, but their average value is 0.2 arcsec for the 14-month and 0.1 arcsec for the 12-month terms. Chandler correctly attributed the 14-month term to free nutation. Euler's period of 10 months was based on a perfectly rigid earth, but the earth's elasticity lengthens the period to 14 months. This term is called the Chandlerian term. The 12-month term is associated with meteorological causes, such as annual variations in atmospheric pressure, vegetation, snow, etc.

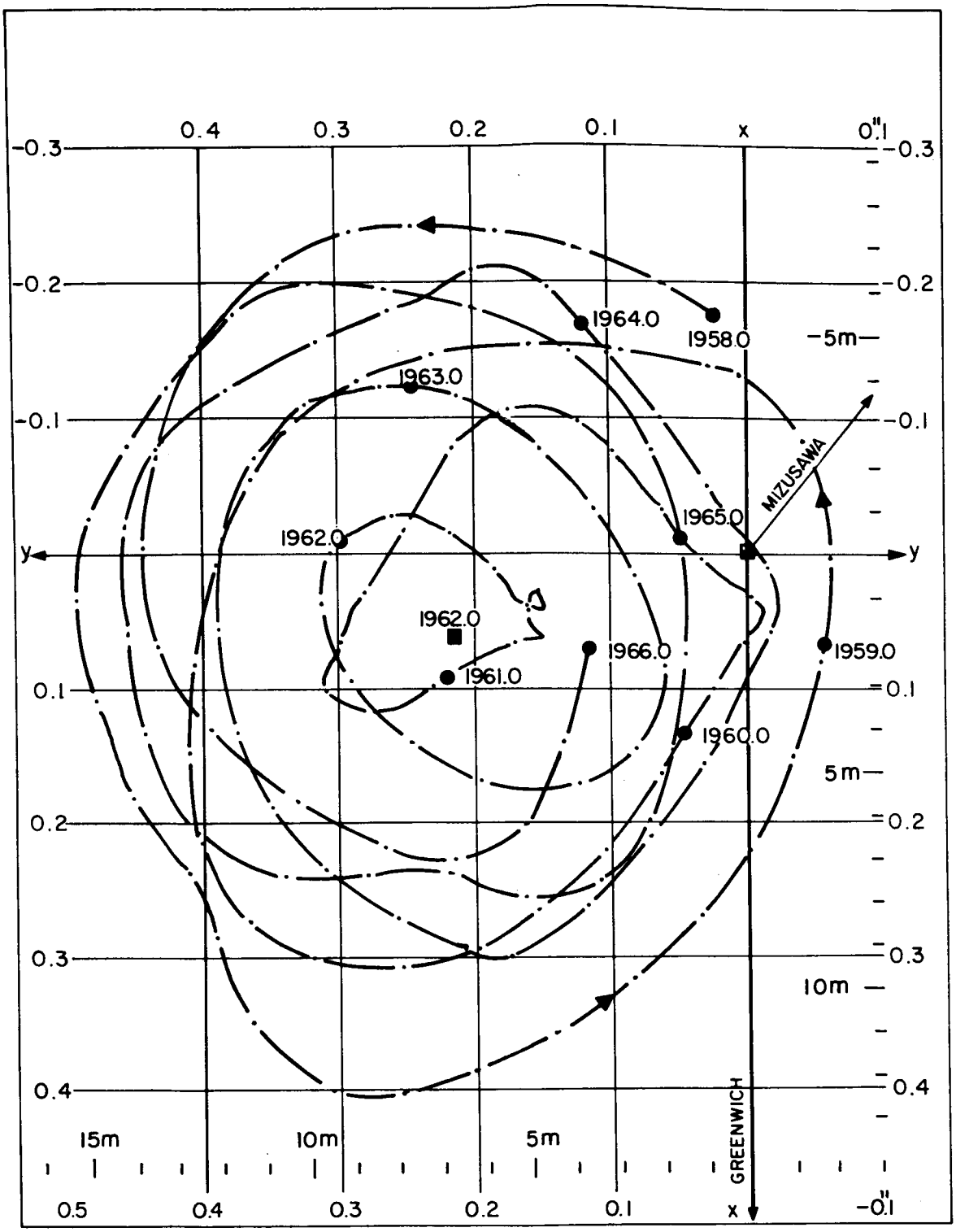


Figure 4. Orbit of instantaneous pole at mean North Pole 1900-1905 (IPMS 1958.0-1966.0).

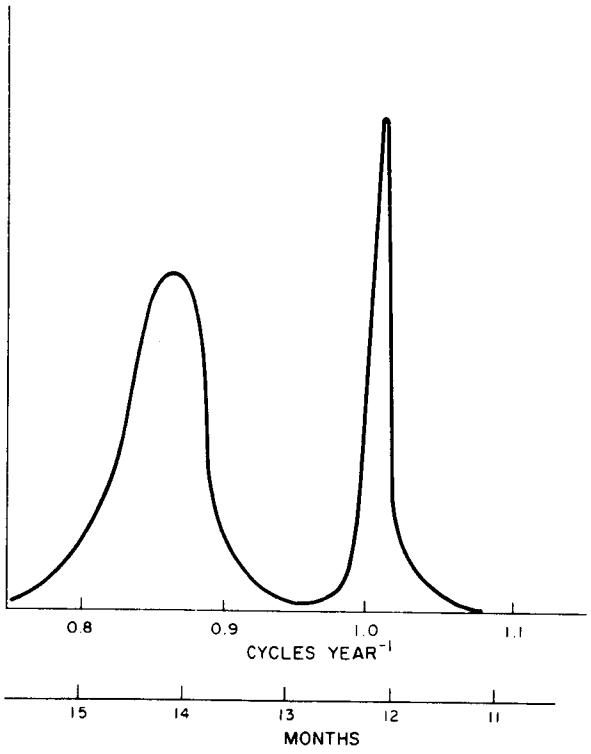


Figure 5. The spectrum of polar motion.

The combination of these two terms shows a variation in amplitude with a period of about 7 years. There is also some indication that there is a secular motion of the mean pole of 0.0032 arcsec year^{-1} (or 10 cm year^{-1}) in the direction of longitude 75° W, but this is not yet fully proved.

The phenomenon of the polar motion is not completely understood nor satisfactorily explained. It must be realized that the existing data are not of the needed accuracy, primarily as far as systematic errors are concerned. In addition, the theoretical work also lacks precision and rigor.

4. THE ROTATION OF THE EARTH

The earth rotates around its spin axis with one revolution per day, more precisely, 1.00273781191 revolutions per solar day. Until about 30 years ago, this angular velocity was assumed to be constant. Actually, it was assumed to be so constant that it was used for centuries for time determination.

With the increase in the accuracy of the observations and the precision of the clocks, it was found that the earth is not actually rotating at a constant velocity: it is affected by 1) a secular variation, 2) irregular variations, and 3) seasonal variations.

On the basis of very old observations of the sun and the moon from records of ancient eclipses, it has been established that the rotation of the earth is continuously slowing down. This change is very small, amounting to an increase in the length of the day of only 1.6 msec century^{-1}. This increase is sufficient to have accumulated almost 3 hours of delay during the last 2000 years. The cause of this deceleration is mainly from tidal frictions.

In the late 1930's, Spencer Jones established also the existence of irregular variations in the rotation of the earth. He found that the observations of the moon, the sun, and the planets did not agree with their ephemerides. Since the error was more than what the observations could allow for, and since it was proportional to the proper motion of the celestial object, he attributed the error to time rather than to position; and since time was based on the rotation of the earth, he concluded that there must exist irregular variations in the rotation of the earth. Figure 6 shows the departures of the observations from theory for the moon. The scale on the right provides, in time, the corresponding accumulated amount of rotation for the earth.

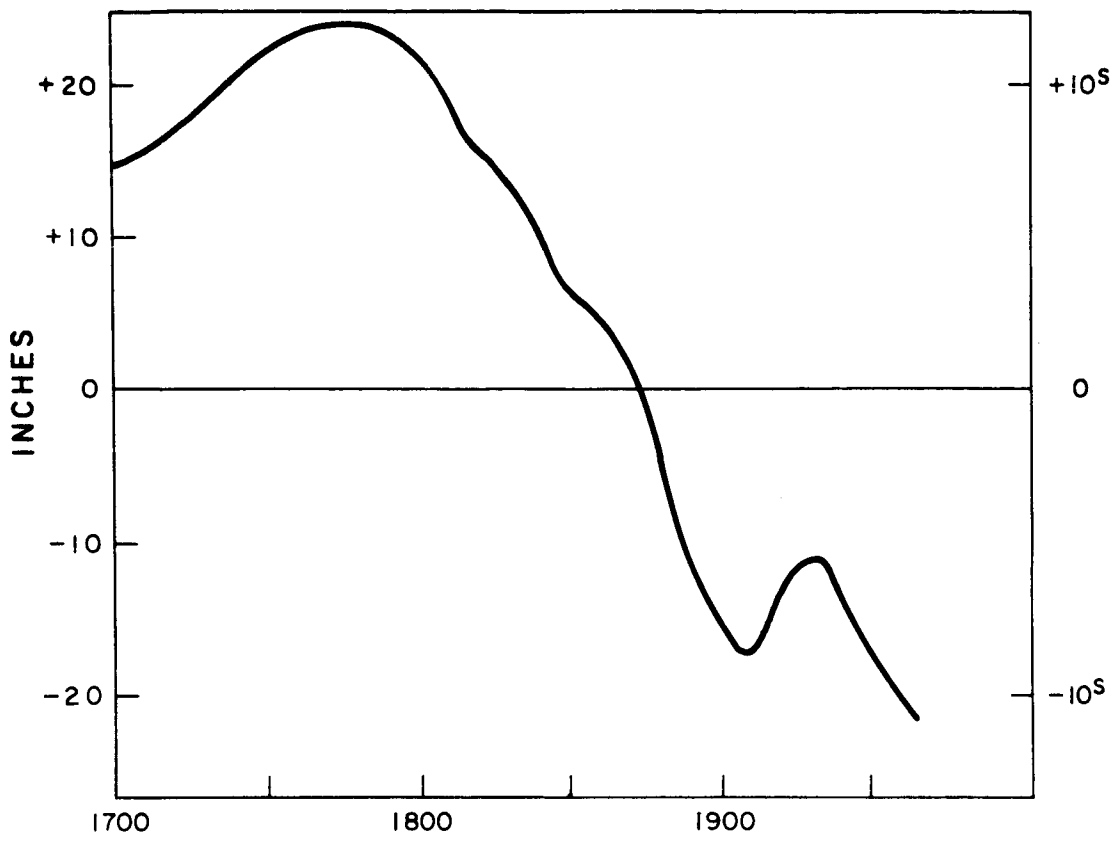


Figure 6. Variations in the moon's longitude.

In 1937, Stoyko showed the existence of a seasonal variation, of about 0.8 msec day^{-1}, in the speed of the rotation of the earth. The earth is slow in the winter and fast in the summer. Figure 7 shows the changes in the speed of rotation from 1958 to 1966.

From the above, it is obvious that the rotation of the earth is not a reliable time-keeping device. Today, time keeping is based on cesium-beam atomic scintillators that are called atomic clocks. These clocks, with a stability of 10^{-10} to 10^{-11} in their frequency, provide a time scale called Atomic Time (AT). It is in AT that we time all phenomena, and since this time is defined by the reading of a clock, it can be assumed to be without error. There is a question whether AT, which is based on phenomena of the microcosmos, is satisfactory for celestial mechanics as well. However, it is too early to answer this question.

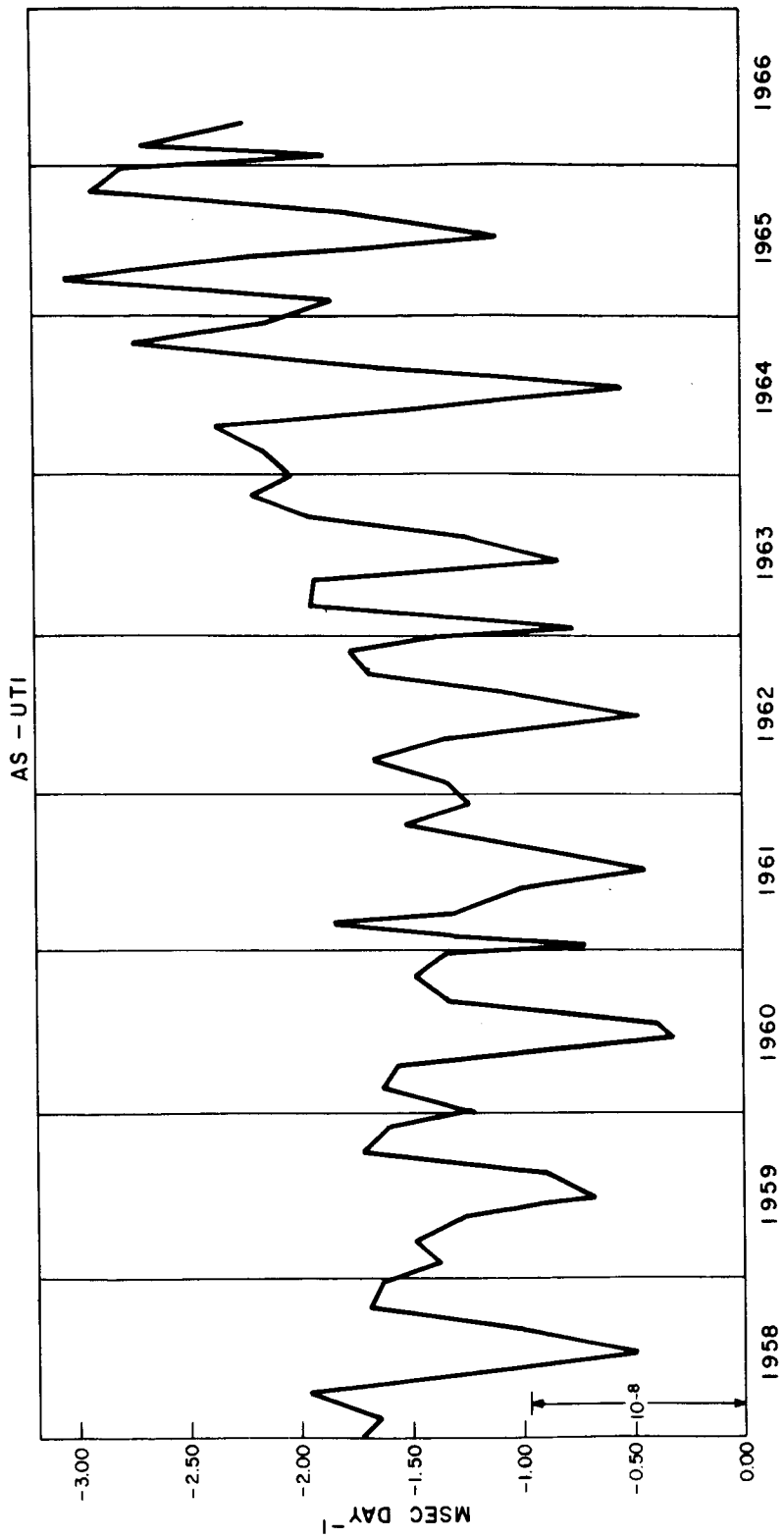


Figure 7. Variations in the speed of rotation of the earth.

The problem, then, is to express the rotation of the earth in terms of AT. And since the rotation of the earth is expressed by the sidereal time, the problem becomes that of finding the relation between sidereal and atomic times.

The fundamental relation between mean sidereal time and mean solar time can be expressed in the form

$$\bar{\theta} = a + bt + ct^2 \quad , \tag{3}$$

where $\bar{\theta}$ is the mean sidereal time and the coefficients a, b, and c, given by Newcomb, have the following values:

$$a = 18^h 38^m 45.^s 836 \quad ,$$

$$b = 8640184.^s 542 \quad ,$$

$$c = 0.^s 0929 \quad .$$

Here t is measured in Julian centuries from noon January 0, 1900.

Since mean sidereal time is observed by timing the meridian transits of stars to a much higher accuracy than solar time can be determined directly from observations of the sun, it is from the solution of equation (3) that solar time is determined. The time obtained with this method is called Universal Time (UT), which after correction for the effect of polar motion (equation (2)), is called UT1. Comparing UT1 with AT, we obtain

$$\Delta U = UT1 - AT \quad . \tag{4}$$

The value of ΔU is found from observations. At a modern observatory, it can be determined with an accuracy of 3 to 5 msec from one night's observations.

From the above, it becomes apparent that to determine $\bar{\theta}$ we must first correct AT by ΔU to obtain UT1 and then solve equation (3). Since ΔU is determined from observations, we need only interpolate for $\bar{\theta}$.

It is also possible to express $\overline{\theta}$ directly as a function of AT. We have

$$\overline{\theta} = a + b(AT) + c(AT)^2 + \Delta\theta \quad , \tag{5}$$

where $\Delta\theta = 1.002738 \cdots \Delta U$.

Currently, the official determination of the rotation of the earth, i. e., of the sidereal time as a function of AT, is made by the Bureau International de l'Heure. Unfortunately, not all the 40 participating observatories produce results of the same quality. The BIH publishes regularly, but with long delay, the necessary corrections to reduce AT in $\overline{\theta}$.

As we define UT from equation (3) based on the rotation of the earth around its axis, we can also define another time based on the rotation of the earth and other planets around the sun and of the moon around the earth. The mean longitude $\overline{\ell}$ of a planet can be expressed in the form

$$\overline{\ell} = A + Bt + Ct^2 \quad .$$

This time is called Ephemeris Time (ET) and can be determined from observations of the moon (with a moon camera) and of the planets and the sun. The name Ephemeris Time comes from the fact that this is the actual Newtonian time definition, on which the ephemerides of the moon and the planets are based.

ET can be compared to UT, and we now have ET - UT = ΔT, which can be determined from observations; but it takes some years before this determination can be made available (see Figure 8). The accuracy with which ET can be determined is about 10 msec. We should expect that ET - AT must be constant, or else there is only a small linear term due to an error in the initial determination of the frequency of the atomic clocks. However, this is not the case, probably because of errors in the determination of ET.

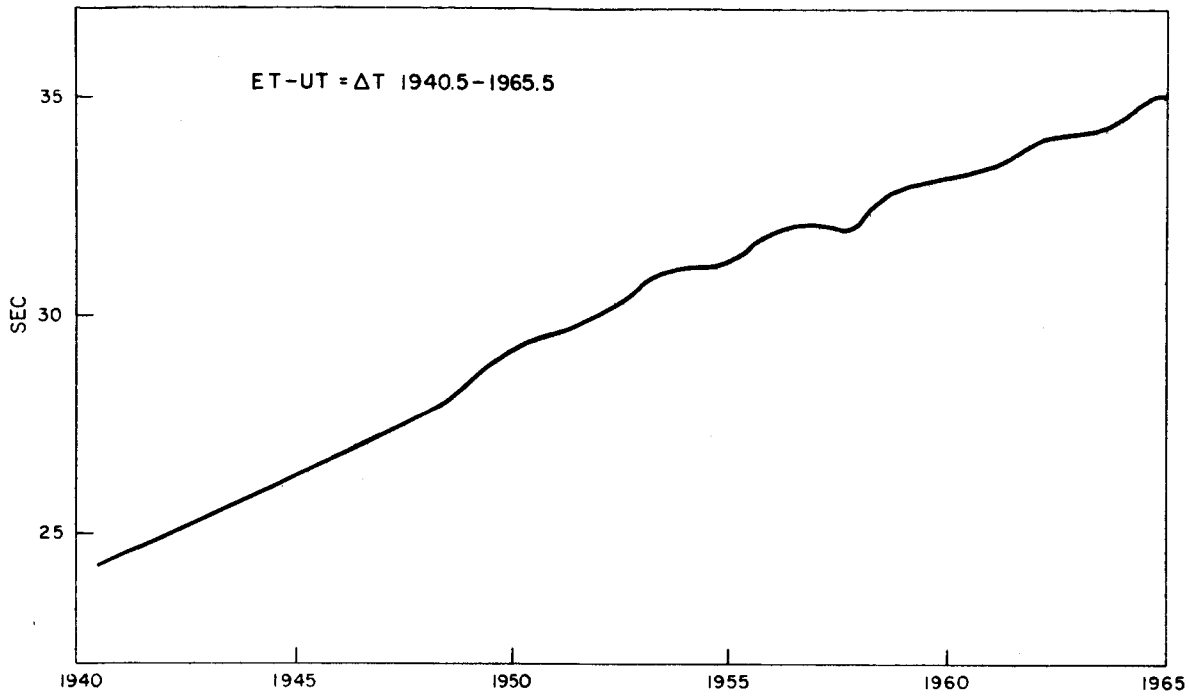


Figure 8. Difference between ET and UT.

Figure 9 shows the geometry of some of the parameters involved with the polar motion and the rotation of the earth.

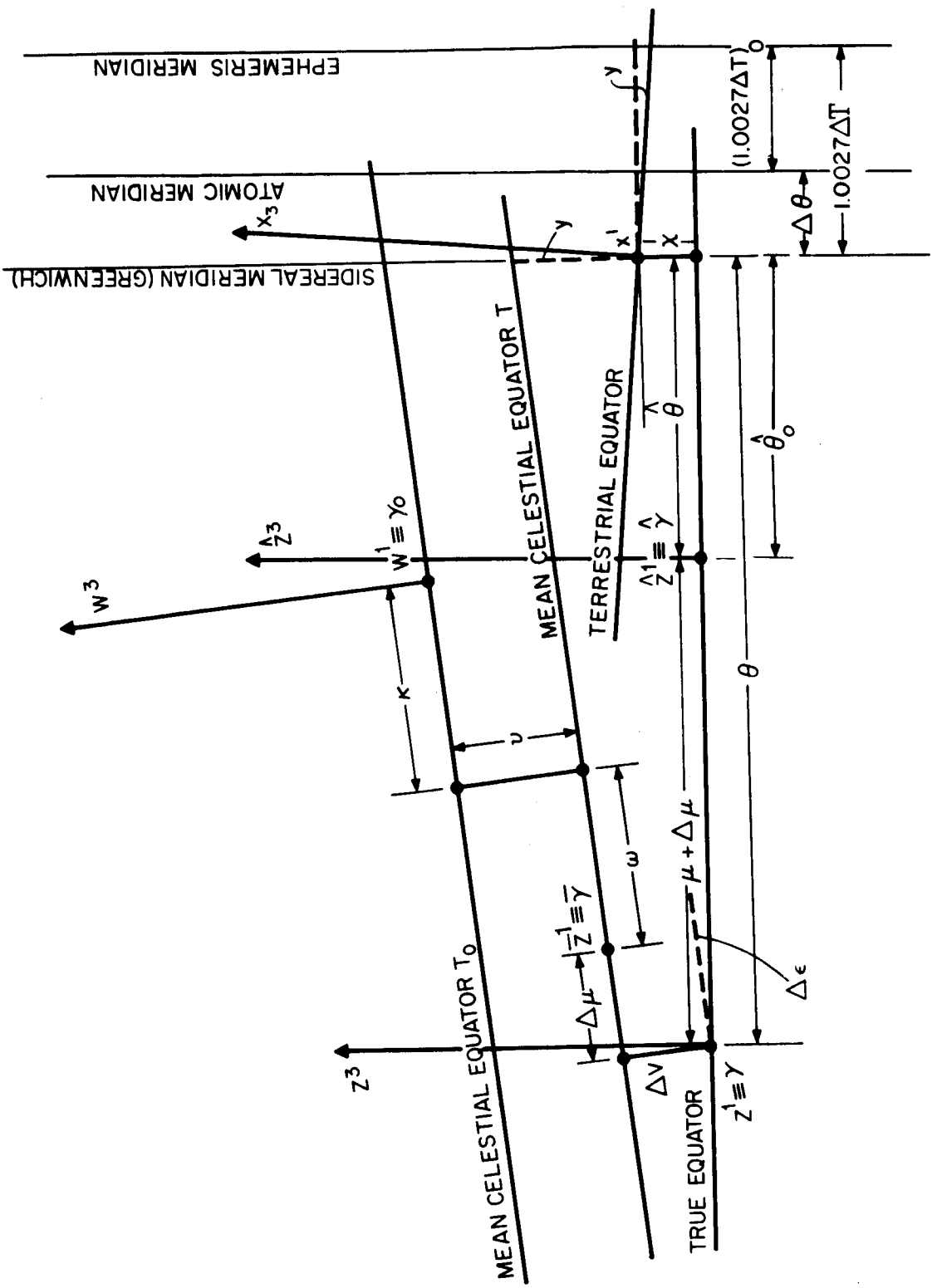


Figure 9. The 1-axis of different reference systems and their relations.

5. GENERAL REMARKS

Two major benefits can be derived from the solution of the problems of polar motion, of the rotation of the earth, and of the determination of time. First is the utilitarian use of the information, primarily for general satellite-orbital determination and analysis, space-probe navigation, guidance, etc. Second is the use of this information as data for geophysical studies related primarily to the physics of the earth's interior.

The present accuracy requirements for both needs are that the data satisfy a positional accuracy of better than 1 m on the surface of the earth. In a few years we will need a 10-cm accuracy. This means that:

A. Timing in the AT system must be provided to ±0.1 msec, corresponding to approximately an 80-cm displacement for an average satellite. This is currently possible with the use of crystal clocks (or, even better, atomic clocks) and VLF receivers. Portable clocks will also be required for the establishment of original time settings.

B. The coordinates x, y of the true instantaneous pole must be given to ±0.02 arcsec (60 cm). The existing ILS data are expected to be accurate to ±0.03 arcsec, but the new IPMS results may be good to ±0.02 arcsec.

C. The true sidereal time, or in other words, the value of ΔU, must be given to ±2 msec, corresponding to 80 cm at $\phi = 30°$. Currently, ΔU is given to ±5 msec (±2 m at $\phi = 30°$), but the ±2-msec accuracy is possible.

D. ET must be determined to ±10 msec, corresponding to a 10-m displacement of the moon. It is a question whether this accuracy is obtainable now; in any case, ET is not made available until after several years.

From the above, it appears as if present techniques are sufficient to provide the required accuracy. It will be an immense problem, since from the 1-m accuracy, we would like to move to the 1-dm accuracy. However,

the real problem is not so much whether this required accuracy can be achieved, but whether the results can be homogeneous and free of systematic errors, and can be made available for use without unreasonable delay, in the adoption of an international reference. Although this is only a question of organization, it is currently a serious problem.

Another problem is to what extent can satellite techniques contribute to the determination of the motion of the spin axis and the rotation of the earth. Since the magnitude of these phenomena is of the order of 10^{-6}, and thus the required accuracy must be of 10^{-7} to have any significance, it looks as if satellite observations can contribute only by the use of special techniques to extract the needed information.

One possibility is to use simultaneous observations from pairs of stations to determine the space direction of the lines connecting them; this will allow the determination of the orientation of the earth's crust. The advantage of this method, compared to the classical one, which uses astronomic techniques, is that we define the body of the earth by means of a number of points on its surface rather than by means of a number of verticals that define directions only. Today, the accuracy that can be reached for the directions of lines from simultaneous observations is quite low, something like 0.5 arcsec. It will be rather unlikely to improve on that accuracy, since in addition to instrumental difficulties, we have the problem of errors in the star catalogs. But with a special program of observing the satellite in the same region of the sky every time and of using a small number of reference stars with precisely determined positions, we could overcome this problem. This method, which is actually the technique used by IPMS and BIH, will improve the accuracy considerably. A synchronous satellite may be an ideal object for this purpose.

Another possibility is to use a pure dynamic method to determine the polar motion. The wandering of the pole should introduce a periodic variation in the J_2 term. With the high accuracy reached now in the determination of J_2 – of the order 10^{-6} – it may be possible in the near future to detect the amount of the polar motion from the introduced perturbation on the satellite motion, and even to determine the values of x and y.

REVIEW OF THE ROTATION OF THE EARTH

E. M. Gaposchkin[1]

1. INTRODUCTION

The study of the rotation of the earth has always been an irresistible attraction for astronomers and geophysicists. For most of history the motion of the earth has been used as the principal timekeeper. Even today, the earth's motion is steady enough to be used for most purposes of time keeping. It was only with the development of precise positional astronomy that time kept in this way became inadequate. What was an annoying problem for the astronomer, which he solved by legislation, became a very fascinating subject for the geophysicist. By studying the motion of the earth, the geophysicist can understand the physical properties of the earth.

It is convenient to separate the motion into three components: 1) precession and nutation, 2) the length of the day (LOD), and 3) the motion of the pole (wobble or Eulerian nutation). To a large extent, this separation is natural, as there are different observations of each and there seem to be different controlling factors. The theory of precession and nutation is formidable (Woolard, 1953) and complete. It hinges on the attraction of the moon, the sun, and the planets on the equatorial bulge; it also is an analytical theory with some constants determined by observation. By definition, therefore, it contains only regular motions, while the LOD and the wobble are observed to have both regular and irregular motions. Precession and nutation seem to be well understood and will not concern us here.

[1]Mathematician, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts.

It is not clear what is meant by regular motion. The emphasis in current work is on statistical inference, spectrum analysis, and discussion in terms of instantaneous frequencies (Munk and MacDonald, 1960, p. 150). From this point of view, there is no "regular" motion, only motion that is random and that that is not. On the other hand, there are some phenomena affecting the motions that have established frequencies.

Studies of the rotation of the earth have been accelerated this century, owing to the discovery of the wobble, the establishment of the International Latitude Service (ILS) and the Time Service, and the development of crystal clocks. Although there has been much pioneering work on the study of the earth's motions since 1950, the fundamentals of these problems still remain to be understood. This fact is appropriately illustrated by an opening remark of Munk and MacDonald (1960, p. 3):

> the 14-month Chandler wobble, the period of which is governed by the ellipticity and rigidity of the Earth; the wobble is generated by random impulses of unknown origin and damped by some unknown imperfection from elasticity, or by some other means.

It is true that new sources of information are becoming available that will bear on this problem. Among others are the exciting prospects of measuring the length of the Paleozoic day by use of fossil coral clocks (Runcorn, 1966; Wells, 1966) and of measuring both the Love numbers pertaining to the solid earth tides and the time-varying gravity field of the earth (Kozai, 1965; Newton, 1965) by use of artificial earth satellites.

2. SUMMARY OF THE SPECTRUM OF EFFECTS

The precession of the earth was known as early as Hipparchus (fl. 160-125). Newton explained the cause of this motion in terms of the attraction by the sun and the moon on the earth's oblateness. Euler formalized this understanding, and in addition showed that a rigid body would have a stable rotation about a principal axis of inertia. Also, he showed that a rigid body like the earth could have a free nutation with a period of A/(C-A) sidereal days. There were many unsuccessful attempts to observe this 10-month period. In 1891 Chandler, a prosperous merchant in Cambridge, Massachusetts, discovered the motion, which bears his name, of a 14-month period. Newcomb immediately showed that the elastic yielding of the earth would lengthen the Eulerian period. Chandler's discovery led to the establishment of the ILS, and it is from about 1900 that these data have become available.

The LOD observations have been made directly for a much shorter period of time, since the establishment of crystal clocks in about 1951.

A direct measurement of the motion of the earth has therefore been made only over a recent interval. This interest in the long-term character of the motion has resulted in the use of lunar-occultation and eclipse records to determine the relatively recent history of the LOD, the use of fossil corals to determine the longer term history, and the use of the magnetic memory of the rocks to measure the paleomagnetic pole and by inference the rotation axis. We summarize in Table 1 whatever information is currently available regarding the earth's rotation. We will not, however, be able to discuss all the quantities presented. The possible causes of the rotation are not generally agreed upon — they are given here only to indicate the kind of thinking that is going on.

Table 1. Summary of the changes in the earth's rotation

Time span	Wobble of the axis			Length of the day		
	Observation	Amount	Cause	Observation	Amount	Cause
2 weeks 1 month	--------- No way of knowing ---------			Crystal clocks astronomic	0.45 msec	Body tides
1/2 year	Latitude service			Crystal clocks astronomic	0.15 msec	Body tides
1 year	Latitude service	12 m	Shift of air mass	Crystal clocks astronomic	1.0 msec 0.025 msec	Winds Body tides
10 years	Latitude service	3 m	Sea-level changes Changes in core	Lunar occultation	1.0 msec	Motion in core
100 years				Lunar occultation	1.0 msec	Sea-level changes
1000 years		<300 m		Ancient eclipses	10 msec	Tidal friction, separation of core from mantle
10^6 years				Fossil clocks		Tidal friction, separation of core from mantle
10^9 years	Paleomagnetism	90°	Shifts of crustal mass, interior changes		unknown	Gravitational sorting out, ocean tides, atmospheric tides

3. THE OBSERVATIONS OF WOBBLE[2]

The ILS essentially measures the angle between the vertical and the pole, by use of known reference stars. The pole, defined by precession — i.e., the angular momentum or instantaneous rotation axis, which differ by 0.003 arcsec (Jung, 1956) — moves with respect to the crust of the earth; hence the term variation of latitude. The ILS provides reduced data points of x and y displacements with respect to an arbitrary reference. The magnitude of the motion is some tens of meters. Figure 1 gives the position of the pole from 1958 to the present.

Unfortunately, the ILS stations have been minimal in number, around five to seven, with only three participating during the entire interval, and those under various administrations. Due to precession, the fundamental reference stars have, of necessity, changed, and many of the problems of treating the astronomical data result from this fact. Munk and MacDonald (1960) note that the irregular changes in the apparent secular motions of the pole are correlated with the change in star catalogs.

The measurement of latitude, or zenith distance, can be made accurate to perhaps 0.1 arcsec. By making 1000 observations, we can improve the reliability of the measurement to less than 0.01 arcsec. This figure (0.01 arcsec) is equivalent to 1.01 feet. The 1000 observations are made over a period of 15 to 20 days. This means that any very short-period variations are averaged out of the data, and we would detect no correlation with any short-lived geophysical event.

[2] Veis (these proceedings) discusses in detail the general statements made here.

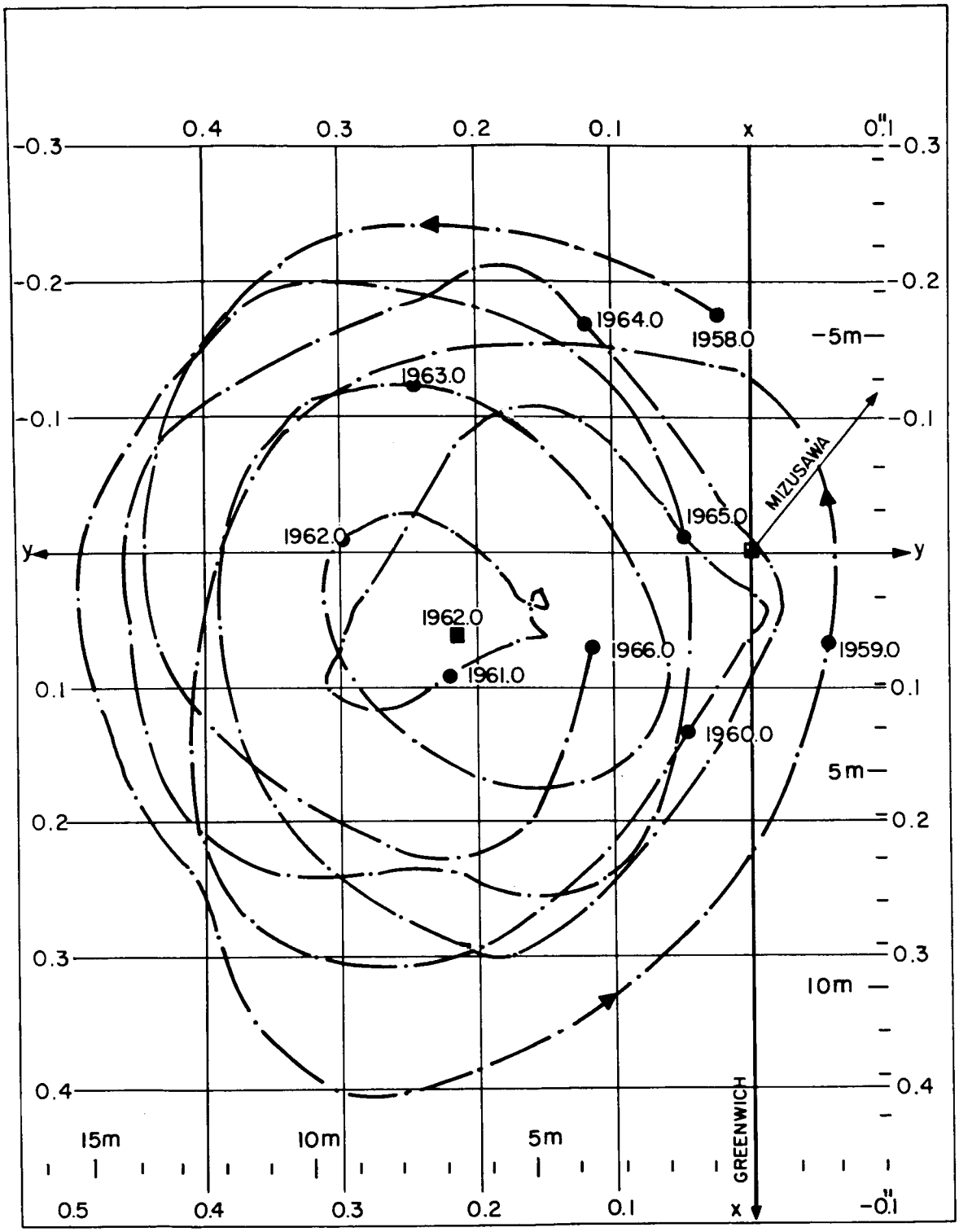


Figure 1. Orbit of instantaneous pole at mean North Pole 1900 to 1905 (IPMS 1958.0 to 1966.0).

Figure 2 (Munk and MacDonald, 1960) gives the ILS data from 1900 to 1960. Figure 3 (Munk and MacDonald, 1960) is a power spectrum analysis of the data, and shows a very pronounced 1-cycle-per-year (cpy) period and a broader 0.85-cpy period. In Figure 2 we have plotted the x and y data with the annual component removed by Fourier analysis. In Figure 4 we give a plot of the annual motion (Munk and MacDonald, 1960) taken from Jeffreys (1962). Jeffreys quotes an uncertainty of about 0.011 arcsec.

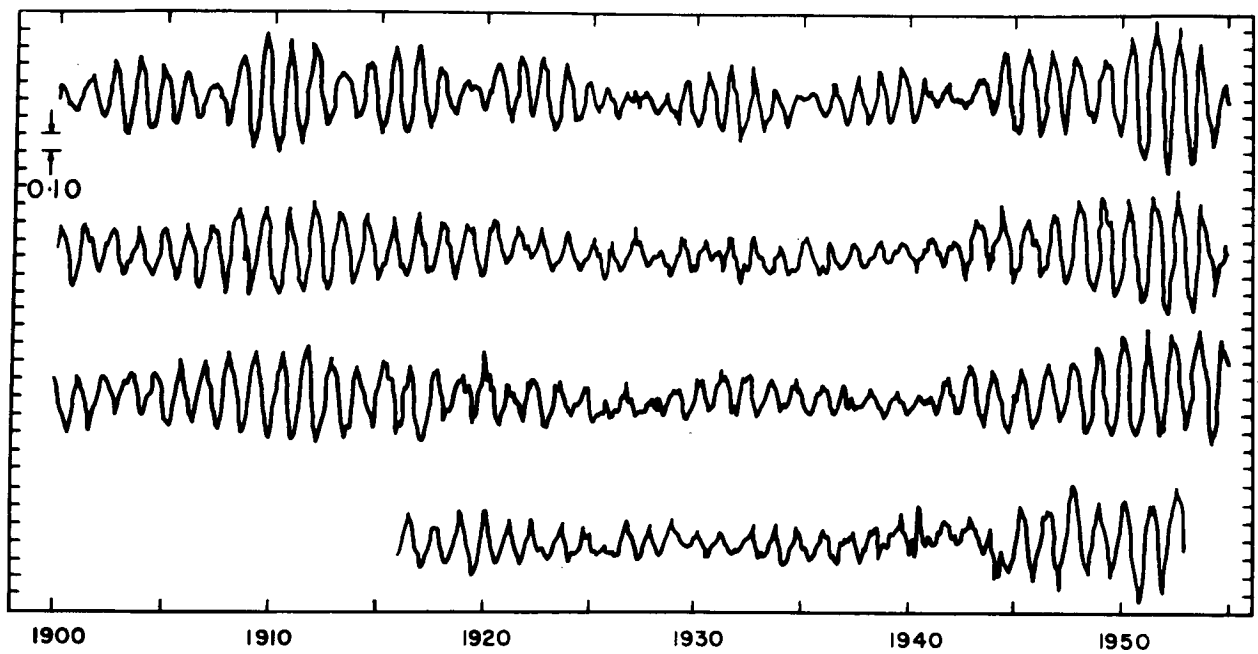


Figure 2. The component m_1 of the unsmoothed ILS observations, before (top) and after (second curve) removal of the seasonal variation; the component $-m_2$ of the unsmoothed ILS observation after removal of the seasonal variation (third curve) and the corresponding nonseasonal variation in the latitude of Washington, as obtained with the photographic zenith tube (bottom). (After Munk and MacDonald, 1960.)

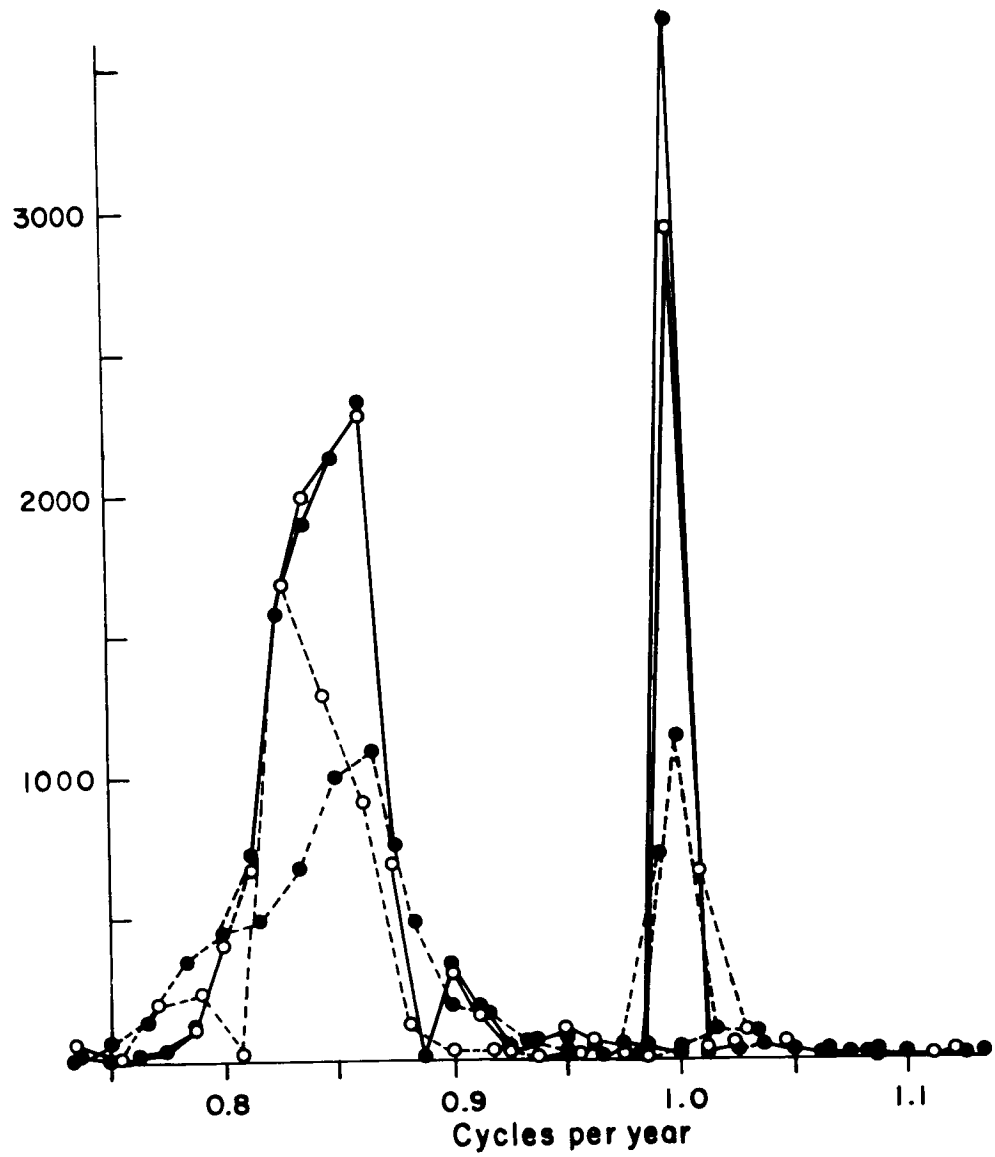


Figure 3. Power spectra of latitude. Solid lines: $\overline{S}_{11}$ and $\overline{S}_{22}$. 1900 to 1954, from ILS. Dashed lines, open circles: $1/2\ S^{+}$, 1891 to 1945, from Kulikov's (1950) complication, according to Rudnick (1956). Dashed lines, solid circles: Washington latitude for 1916 to 1952. The ordinate gives power density in units of $(0.01\ \text{arcsec})^2$ cpy. (After Munk and MacDonald, 1960.)

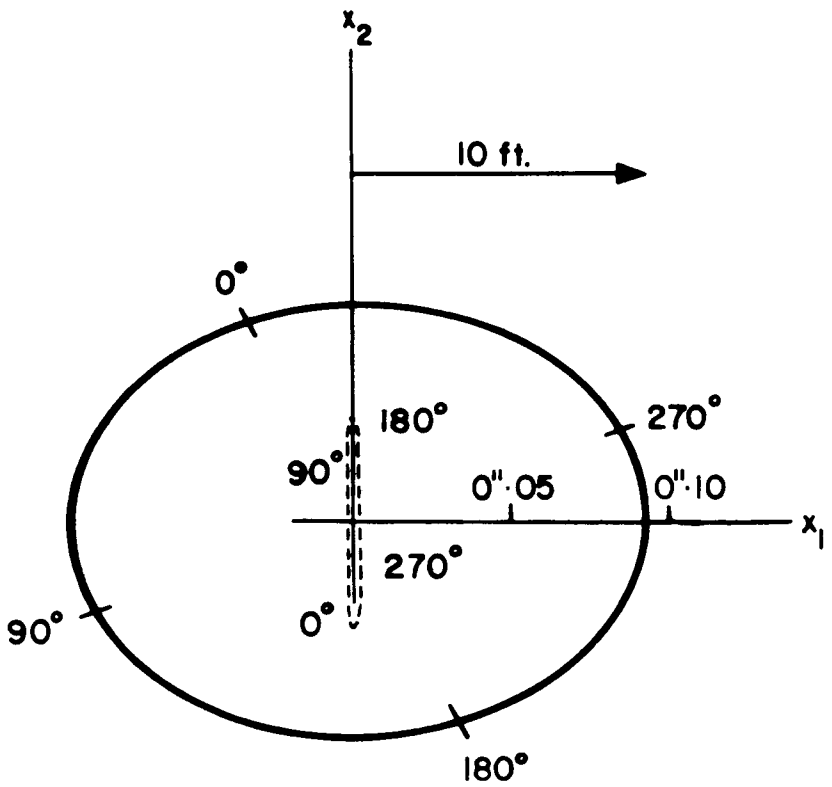


Figure 4. The annual motion of the poles of rotation (solid) and excitation (dashed) with respect to the reference system centered at the mean pole, with x_1 toward Greenwich, x_2 toward 90° east of Greenwich. The positions of the pole at $\odot = 0°$ (January 1), $\odot = 90°$ (April 2), $\odot = 180°$, and $\odot = 270°$ are indicated. The amplitude scale in arcsec is given along the x_1 axis; 0.01 arcsec is equivalent to 1 foot of polar displacement. (After Munk and MacDonald, 1960.)

4. FORMALISM

Following Munk and MacDonald (1960), we can write the Eulerian equations of motion in a moving coordinate system by

$$L_i = \frac{dH_i}{dT} + \epsilon_{ijk}\,\omega_j\,H_k \quad , \tag{1}$$

where L is the applied torque and H the angular momentum. It is permissible to separate the angular momentum into two parts, the first associated with the rotation around some axis (e.g., the solid earth) and the second associated with auxiliary motions with respect to this system (e.g., motion of air mass):

$$H_i = C_{ij}(t)\,\omega_j + h_i(t) \quad , \tag{2}$$

$$C_{ij} = \int_V \rho\left(x_k x_k\,\delta_{ij} - x_i x_j\right) dV \quad , \tag{3}$$

$$h_i = \int_V \rho\,\epsilon_{ijk}\,x_i\,u_k\,dV \quad . \tag{4}$$

This separation is quite arbitrary and is chosen for convenience. Substituting equations (2), (3), and (4) into (1) gives

$$L_i = \frac{d}{dt}\left(C_{ij}\,\omega_j + h_i\right) + \epsilon_{ijk}\,\omega_j\left(C_{kl} + h_k\right) \quad . \tag{5}$$

We can now specialize to a chosen reference system. We choose x_i fixed on the solid earth.

For treating the motion of the pole, assuming the motion to be small, let

$$C_{11} = A + c_{11} \quad , \quad C_{12} = c_{12} \quad , \quad C_{13} = c_{13} \quad ,$$

$$C_{22} = A + c_{22} \quad , \quad C_{23} = c_{23} \quad , \quad C_{33} = C + c_{33} \quad , \tag{6}$$

where A, A, and C are the principal moments of inertia. Let

$$\omega_1 = \Omega m_1 \quad ,$$

$$\omega_2 = \Omega m_2 \quad ,$$

and

$$\omega_3 = \Omega (1 + m_3) \quad , \tag{7}$$

where Ω is the rotation speed of the earth ($\Omega = 0.729 \times 10^{-4}$ rad $\sec^{-1}$). Therefore, c_{ij}/C, m_i, and h_i/C are small dimensionless parameters whose products can be ignored. The Liouville equation reduces to

$$\frac{\dot{m}_1}{\sigma_r} + m_2 = \phi_2 \quad , \tag{8}$$

$$\frac{\dot{m}_2}{\sigma_r} - m_1 = \phi_1 \quad , \tag{9}$$

$$\dot{m}_3 = \dot{\phi}_3 \quad , \tag{10}$$

$$\sigma_r = \frac{C - A}{A}\,\Omega \quad , \tag{11}$$

$$\Omega^2 (C - A)\,\phi_1 = \Omega^2 C_{13} + \Omega \dot{C}_{23} + \Omega h_1 - \dot{h}_2 - L_2 \quad , \tag{12}$$

$$\Omega^2 (C - A)\,\phi_2 = \Omega^2 C_{23} - \Omega \dot{C}_{13} + \Omega h_2 - \dot{h}_1 + L_1 \quad , \tag{13}$$

$$\Omega^2 C\,\phi_3 = -\Omega^2 C_{33} - \Omega h_3 + \Omega \int_0^t L_3\, dt \quad . \tag{14}$$

The first two expressions pertain to the wobble, and the third to the LOD. Since we have a linear system, we can compute them:

$$m = m_1 + i\, m_2 \quad , \qquad \phi = \phi_1 + i\, \phi_2 \quad , \qquad i = \sqrt{-1} \quad . \tag{15}$$

Therefore,

$$i \frac{\dot{m}}{\sigma_r} + m = \phi \quad . \tag{16}$$

This form of the equations nicely separates the two effects into the LOD and the polar wobble. By writing the equation in this way, we have expressed on the left-hand side the observed astronomical quantities and on the right-hand side the postulated geophysical quantities. The geophysical problem, then, is to evaluate the right-hand side. The right-hand side, ϕ, is called the excitation function. A new source of information from artificial earth satellites is the determination of C_{ij}.

Equations (12), (13), and (14) are good for computing the excitation function when changes in angular momentum are well separated from changes in the products of inertia. However, to separate the changes

in redistribution of matter from those of relative motion, which enter both c_{ij} and h_i, Munk and MacDonald (1960) and Munk and Groves (1952) give

$$\Omega^2 (C - A)\,\phi = \int_V \Delta\rho\, F^m\, dV + \int_V \rho\, F^v\, dV + F^t \quad ,$$

and

$$\Omega^2 (C - A)\,\phi_3 = \int_V \Delta\rho\, F_3^m\, dV + \int_V \rho\, F_3^v\, dV + F_3^t \quad , \tag{17}$$

where dV is the element of volume. In spherical coordinates,

$$dV = r^2 \sin\theta\, dr\, d\theta\, d\lambda \quad , \tag{18}$$

where r is the radial distance, θ the colatitude, and λ the longitude. For mass transport,

$$F^m = -r^2 \Omega^2 \sin\theta \cos\theta\, e^{i\lambda} \quad ,$$

and

$$F_3^m = -r^2 \Omega^2 \sin\theta \quad . \tag{19}$$

For velocity,

$$F^v = \left\{ -2\Omega r \cos\theta \left[u_\lambda - i\,(u_\theta \cos\theta + u_r \sin\theta) \right] + r\,(\dot{u}_\theta + i\,\dot{u}_\lambda \cos\theta) \right\} e^{i\lambda} \quad ,$$

and

$$F_3^v = -\Omega r \sin\theta\, u_\lambda \quad , \tag{20}$$

where u_λ, u_θ, and u_r are the components of the velocity in the λ, θ, and r directions. For torque,

$$F^t = -L_2 + i L_1 \quad ,$$

and

$$F_3^t = \Omega \int_0^t L_3 \, dt \quad , \tag{21}$$

$$L_i = \int_V \rho \epsilon_{ijk} x_j f_k \, dV + \int_S \epsilon_{ijk} x_j p_{km} n_m \, dS \quad , \tag{22}$$

where the f_k are the body forces and the p_{km} are surface stresses.

The earth is not rigid, and we cannot directly apply the above Liouville equations. In principle, we must solve the very complicated problem of elasticity. The question of the proper model for this problem is even today unresolved, as is the solution of this problem. We can conveniently avoid this problem by introducing the appropriate Love numbers.

To define the Love numbers we consider the earth under a disturbing potential of second-degree U (Jeffreys, 1962; Love, 1911). The resulting deformation defines the Love numbers as follows: the number h, the ground lifted by $(h/g)\, U_{surface}$; the number k, the additional gravitational potential at the displaced surface arising solely from the redistribution of mass, $(k/g)\, U_{surface}$; and the number ℓ, the horizontal displacement of the surface with components $(\ell/g)\, \partial U/\partial\theta$ and $(\ell/g)(1/\sin\theta)(dU/d\lambda)$.

[The notation here is reasonably uniform; however, there are the following variants: If we let S_n be a spherical surface harmonic of degree n, then $r^n S_n$ and $r^{-n-1} S_n$ are solid spherical harmonics of degree n. In general, the potential deforming the earth will be

$$U = \sum_{n=2}^{\infty} a_n \left(\frac{r}{a_e}\right)^n S_n \quad .$$

This deformation is such that the ground is lifted by

$$\frac{1}{g}\sum_{n=2}^{\infty} h_n a_n \left(\frac{r}{a_e}\right)^n S_n \Bigg|_{r=a_e} .$$

Here h_n is the generalized Love number. Similarly, the incremental potential at the surface can be expressed as

$$\Delta g = \frac{1}{g}\sum_{n=2}^{\infty} k_n a_n S_n$$

and the external incremental potential must be

$$\Delta g = \frac{1}{g}\sum_{n=2}^{\infty} k_n \left(\frac{a_e}{r}\right)^{n+1} a_n S_n .$$

Assuming a homogeneous incompressible sphere, we can compute h_n, k_n, and ℓ_n from the rigidity (Munk and MacDonald, 1960, p. 29).

[For most purposes, the discussion is limited to n = 2. Therefore, Jeffreys refers to k_2 as k and denotes a_2/a_e^2 by k_2. On the other hand, Kozai gives explicit forms for a_2 and a_3:

$$a_2 = \frac{GM_\ell (a_e)^2}{r_\ell^3}$$

$$a_3 = \frac{GM_\ell (a_e)^3}{r_\ell^4}$$

Kozai's k_2 and k_3 are the same as in the expression for Δg above.]

In addition, we can evaluate the Love number k for a fluid earth, given the ellipticity, and the Love number k appropriate to the secular motion (Munk and MacDonald, 1960, p. 26). Both determinations give the same value, $k_f = 0.96$.

4.1 The Free Wobble

Equation (16) for a rigid earth predicts a period of $\sigma_r = [(C - A)/A]\Omega$ with $\phi = 0$. The motion of the pole m causes a readjustment of the elastic earth in terms of its moments of inertia, as given in the following equations.

By using the definitions of the Love numbers, we can express the effect of the motion of the pole m on the inertia tensor C_{ij}, through deformation owing to rotation,

$$C_{ij} = \frac{k}{k_f} \frac{\omega_i \omega_j}{\Omega^2} (C - A) \quad , \qquad i \neq j \quad . \tag{24}$$

If we write this out for the components c_{13}, c_{23} and use the above definitions, then

$$c_{13} = (C - A) \frac{k}{k_f} m_1 \qquad \text{and} \qquad c_{23} = (C - A) \frac{k}{k_f} m_2 \quad . \tag{25}$$

Using equations (25) and (12), (13), and (14), we compute

$$\phi_e = \frac{k}{k_f} \left(m - \frac{i}{\Omega} \dot{m} \right) \quad , \tag{26}$$

which, inserted into equation (16), gives

$$i \frac{\dot{m}}{\sigma_0} + m = \phi' \quad , \qquad \phi' = \phi \frac{k_f}{k_f - k} \quad , \tag{27}$$

$$\frac{\sigma_0}{\sigma_r} = \frac{k_f - k}{k_f} \quad , \tag{28}$$

assuming $(1/\sigma_r) + (k/k_r \Omega) \approx 1/\sigma_r$. This approximation is good to 0.1%.

It is customary at this point to note the period of the free nutation of the earth, 14 months, and to use this to compute the Love number k. From this calculation, the rigidity of the earth can be estimated. The Love number k agrees reasonably well with results from seismic evidence, and has a value of 0.29.

This formulation in terms of Love numbers has a very great advantage. Since the event that leads to an excitation function can load the earth, the event will have a second effect on the dynamics of the earth. The approach is to compute the excitation function as if the earth were rigid, as in the case of the elasticity of the earth, which results in an additional excitation function. Therefore, the total excitation function is the sum of these, and can be combined into

$$\phi = \phi + \frac{k}{k_f} m \tag{29a}$$

when the event does not load the earth, and into

$$\phi = \phi + \frac{k}{k_f} m + k' \frac{k}{k_f} m \tag{29b}$$

when it does.

If we now define a transfer function

$$K = \frac{k_f}{k_f - k} \tag{30a}$$

when the event loads the earth, and

$$K = (1 + k') \frac{k_f}{k_f - k} \tag{30b}$$

when it does not, we can then write equation (27) as

$$\dot{m} = i(\sigma_0)(m - \psi) \quad , \tag{31}$$

where $\sigma_0 = \sigma_r[(k_f - (k/k_f)]$ and ψ is $K\phi$. We can therefore proceed to evaluate ϕ as if the earth were rigid, solving the equation with the appropriate value of K from Table 2.

Table 2. The transfer function K (from Munk and MacDonald, 1960, p. 43)

	no load	load
wobble	1.43	1.00
LOD	1.00	0.70

5. BODY TIDES AND THEIR EFFECTS ON LATITUDE DETERMINATION AND THE LOD

The solid-earth tidal deformation and consequent motion cause the effective vertical of the earth to change. These changes must be known for us to be able to reduce the observations to a mean earth. The size of the effect is small and is comparable to the motion itself. In addition, the effect is periodic in such a way that it can disguise another motion.

As an example let us illustrate the size of the phenomenon and the method of calculating it. The vertical, to which latitude is referred, is affected by the tide as shown in Figure 5.

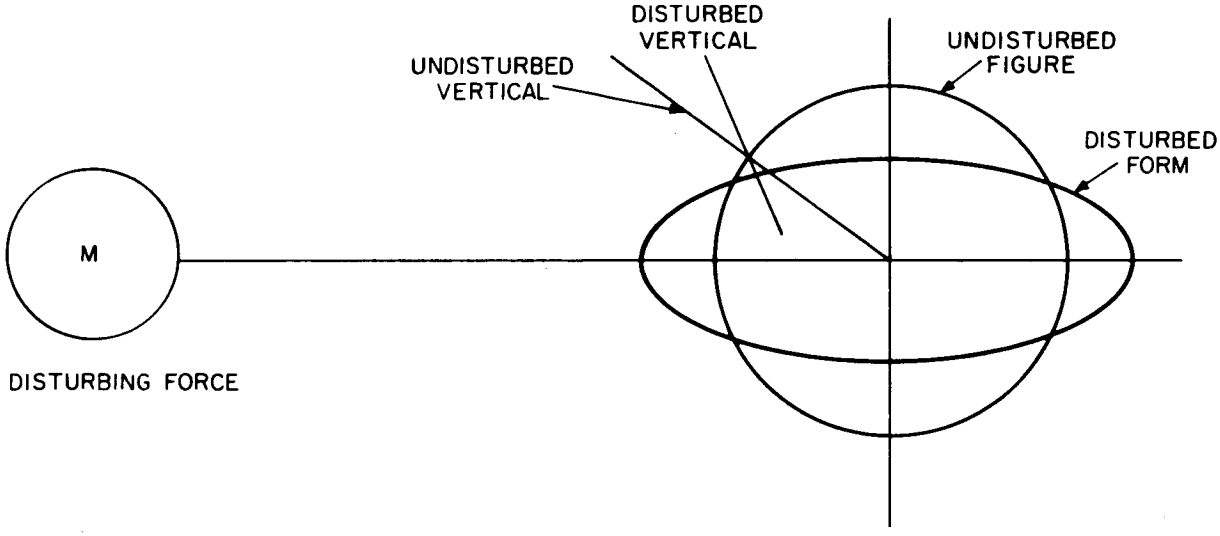


Figure 5. Change in the vertical due to tidal deformation.

We can express the change in vertical toward the south as

$$\delta\theta = -\frac{1}{g\,a_e}\frac{\partial U}{\partial \theta} \qquad \text{direct tidal force,}$$

$$-\frac{k}{g\,a_e}\frac{\partial U}{\partial \theta} \qquad \text{tidal bulge,}$$

$$+\frac{\ell}{g\,a_e}\frac{\partial U}{\partial \theta} \qquad \text{motion of the site southward.}$$

If we adopt $1 + k - \ell = 1.0 + 0.29 - 0.07 = 1.22$, the deflection is away from the tide-producing body.

We can express the tide due to the moon as

$$U = g\,K_{(}\,b\;P_n^m(\theta)\cos m\beta(\lambda, t) \quad ,$$

where P_n^m are the associated Legendre functions and

$$K_{(} = \frac{3}{2}\frac{m_{(}}{m}\frac{a^3}{r_{(}^3}\,a = 53.7 \text{ cm} \quad .$$

Rationalizing several constants, we can express the deflection as

$$\delta\theta = 0\overset{''}{.}0212\;b\,P_n^m(\theta)\cos m\beta(\lambda, t) \quad ;$$

the constants b are taken from Table 7.1 of Munk and MacDonald (1960), reproduced here as Table 3. Added to their table is the total amplitude, under the heading K″. Note that the largest term,

Table 3. Parameters for some equilibrium tides (from Munk and MacDonald, 1960, p. 69)

Species	Symbol	Period	b	K″	f(θ)	β(λ, t)
Long period	Lunar	18.6 years	0.066	0.0014	$3/4(1/3 - \cos^2\theta)$	$(N_☾ - N_{☾0})$
	Sa	1 year	0.012	0.00025		$(☉ - 1.°8)$
	Ssa	1/2 year	0.073	0.0015		$2(☉ - 79.°8)$
	MSm	$31^d.85$	0.016	0.00034		$(☾ - ☉ + p_☾)$
	Mm	$27^d.55$	0.083	0.0018		$(☾ - p_☾)$
	MSf	$14^d.77$	0.014	0.00030		$2(☾ - ☉ + 79.°8)$
	Mf	$13^d.66$	0.156	0.0033		$2☾$
	...	$13^d.63$	0.065	0.0014		$2☾ + N_☽ - N_{☾0}$
Diurnal	O_1	$25^h.82$	0.377	0.00798	$\sin\theta\cos\theta$	$(qt + ☉ - 2☾ - 169.°8 + \lambda)$
	P_1	$24^h.07$	0.176	0.00373		$(qt - ☉ - 10.°2 + \lambda)$
	K_1	$23^h.93$	0.531	0.01122		$(qt + ☉ + 10.°2 + \lambda)$
	N_2	$12^h.66$	0.174	0.00369	$1/2\sin^2\theta$	$2(qt + ☉ - (3/2)☾ + (1/2)p_☾ - 79.°8 + \lambda)$
Semidiurnal	M_2	$12^h.42$	0.908	0.0192		$2(qt + ☉ - ☾ - 79.°8 + \lambda)$
	S_2	$12^h.00$	0.423	0.00895		$2qt$
	K_2	$11^h.97$	0.115	0.00242		$2(qt + ☉ - 79.°8 + \lambda)$

☉ is the longitude of the "mean sun," increasing by 0.°0411 per mean solar hour.
☾ is the mean longitude of the moon, increasing by 0.°5490 per mean solar hour.
$p_☾$ is the mean longitude of lunar perigee, increasing by 0.°0046 per mean solar hour.
$N_☾$ is the mean longitude of the lunar ascending nodes, increasing by -0.°0022 per mean solar hour.
q is the angular velocity of the earth relative to the mean sun, 15° per mean solar hour.
$\Omega = q + d☉/dt$ is the angular velocity of the earth relative to the stars, 15.°0411 per mean solar hour.

$$\delta\theta = 0\overset{''}{.}0192 \sin\theta \cos\theta \cos 2(qt + \odot - \mathbb{C} - 79\overset{\circ}{.}8 + \lambda) \quad ,$$

has a 12.42-hour period. In comparison, the annual term is

$$\delta\theta = 0\overset{''}{.}0003 \sin 6 \cos\theta$$

and the semiannual term is

$$\delta\theta = 0\overset{''}{.}0023 \cos\theta \sin\theta \quad .$$

This illustrates a complication in analyzing the data. The reference stars shift slowly, owing to annual precession. Therefore, since the observations are made with respect to the same reference stars each night, the 12-hour term could be mistaken for the annual term.

These tides limit the accuracy with which the LOD can be measured, i.e., UT1 determined. Consider that the motion of the pole m will displace the meridian. The displacement relative to the stars in the westerly direction is $\cos\theta\ \Re(ime^{i\lambda})$, where $\Re$ is the real part and 1 sec of time equals 15 sin θ arcsec. With m in sec of time, the stars are late and the earth is slow, by

$$\tau = \frac{1}{15} \cot\theta\, \Re(ime^{i\lambda}) \quad ;$$

i.e.,

$$\Delta_{LOD} = \frac{\partial\tau}{\partial t} (LOD) \quad .$$

Now let $m = M e^{i\sigma t}$ (the annual term); then,

$$\tau = \frac{1}{15} \cot \theta \, \mathcal{R}\left[i M e^{i(\sigma t + \lambda)}\right]$$

$$\frac{d\tau}{dt} = -\frac{1}{15} \cot \theta \, \mathcal{R}\left[M \sigma e^{i(\sigma t + \lambda)}\right] \quad .$$

It is observed that M = 0.084 arcsec,

$$\sigma = \frac{2\pi}{365.25} \quad ,$$

$$\tau = 5.6 \cot \theta \text{ msec} \quad ,$$

$$\Delta_{LOD} = 0.096 \cot \theta \text{ msec} \quad ,$$

for the amplitudes of the annual variation. At $\theta = \pi/4$, this is one-fourth the observed annual variation.

As in the determination of latitude, the body tide has an effect on the determination of the LOD. As for the vertical, we have

$$\xi = \frac{1 + k - \ell}{g a_e \sin \theta} \frac{\partial U}{\partial \lambda} \quad ,$$

$$\tau = \frac{\xi}{2\pi} \quad .$$

If we take b = 0.56 and U as before, we get for the annual effect

$$\xi = 0''.0119 \cos\theta \cos\sigma t \quad ,$$

$$\tau = 0.8 \cot\theta \cos\sigma t \text{ msec} \quad ,$$

or

$$\Delta_{LOD} = \frac{d\tau}{dt} \times LOD = 0.0114 \cot\theta \sin\sigma t \text{ msec} \quad .$$

Errors in LOD due to an uncertainty in the wobble, to the east-west deflection of the vertical by tides, and to zenith refraction are all of the order of 1 msec. These errors are comparable to the accuracy of measurements achieved by modern methods. Therefore, it is finally the lack of knowledge concerning these geophysical events that limits the accuracy of time keeping and of measuring the wobble.

6. THE VARIATION OF THE POLE AND ITS EXPLANATION IN TERMS OF WEATHER

6.1 The Annual Variation

The free nutation of the pole is "explained" in terms of the elasticity of the earth. The other motions must be explained in terms of excitation forces, because the earth must be a dissipative system and because, for motions to continue, there must be a driving force. We must then examine the motion for well-defined effects and ascribe driving forces. Figure 3 shows a 12-month and a 14-month period. The 12-month period is very clearly present and quite sharply defined. The 14-month period is less sharply defined, indicating a broader band of excitation.

Since Spitaler (1897, 1901) and Jeffreys (1916), it has been conjectured that the annual motion of the pole could be controlled by the atmospheric motions. It has also been conjectured that the motions of the atmosphere, those not strictly annual, could be the driving force of the 14-month variation. Both the conjectures are reasonable, in that the observed magnitudes of the variation could account for the motion, but until now analysis of the observed atmospheric data has not borne these theories out.

To a large extent the atmospheric motions are governed by geostrophy. This means that the pressure gradiant is controlled by the Coriolis terms in the Naviar Stokes equations. From any text on dynamic meteorology we can write the velocities in terms of the pressure gradient:

$$\rho U_\lambda = \frac{1}{2a_e \Omega \cos\theta} \frac{\partial P}{\partial \theta} \quad , \qquad \rho U_\theta = - \frac{1}{2a_e \cos\theta \sin\theta} \frac{\partial P}{\partial \lambda} \quad . \qquad (32)$$

We can use equations (32) in equations (17) to evaluate the excitation function due to the momentum of the atmosphere:

$$\phi = -\frac{a^2}{\Omega^2 (C-A)} \int_V \left[\frac{d}{d\theta} \left(P \sin\theta e^{i\lambda} \right) + i \frac{d}{d\lambda} \left(P \cos\theta e^{i\lambda} \right) \right] d\theta \, d\lambda \, dr$$

$$\equiv 0 \quad .$$

This result would enable us virtually to ignore the momentum changes of the atmosphere as a factor in the motion of the pole.

It would be very interesting to investigate the effect of the departures from geostrophy on ϕ. Munk and MacDonald (1960, p. 105, 106) derive the result

$$\gamma \sim \frac{\rho F_3^v}{\Delta\rho \, F_3^m} \quad .$$

Equations (19) and (20) show that $|F^v| \approx |F_3^v|$ and $|F^m| \approx |F_3^m|$. The ageostrophic effect is smaller than the geostrophic effect by the order of the Rossby number R_0. Therefore,

$$\frac{\text{the effect of ageostrophic motion}}{\text{the effect of mass distribution}} \cong \gamma R_0 \quad .$$

Now $R_0 \cong 0.1$ and $\gamma >> 1.0$ for the atmosphere owing to the thermal-wind relations. The size of γR_0 indicates that the momentum in the atmosphere may be important in controlling the motion of the pole. Of course, the momentum is accepted as a major factor in controlling the LOD. The question of ageostrophic motion is not closed, and must await actual measurements of the momentum of the atmosphere, as velocities inferred from pressure measurements and from equations (32) have no effect.

In addition, the theory of geostrophic motion has a singularity at the equator; the integral for the excitation function does not. However, the observed motion is still apparently geostrophic. This is an area of continued theoretical interest.

In principle, the integration of equation (17) is carried out along the geoid. The limits of integration must be modified to correspond with departures from this reference surface, i.e., mountains, etc.

The mass distribution of the atmosphere can be approached only through direct observations of the pressure and the density or the temperature of the atmosphere. There have been many determinations of the atmospheric excitation, the most recent by Munk and Hassen (1961).

The evaluation, in a practical sense, of the excitation function depends on the fact that the oceans as a whole react as an inverted barometer. Munk and MacDonald (1960, p. 100) demonstrate this fact with some assumptions that seem valid. Groves (1957) presents observations indicating that sea-level response is roughly two-thirds of the inverted barometer. There is certainly a question of its validity near coast lines, and the departures of this law from observation should be examined. The utility of this law is that the ocean redistributes itself for pressure variations, to annul horizontal gradients on the sea bottom. The bottom pressure varies with time because of the variation of the total atmosphere over the oceans. Thus, we need only land-based observations to compute the excitation function for the whole earth. The most recent determination by Munk and Hassen (1961), using meteorological observations from 1873 to 1950, gives:

$$\phi_1 = -1.8 \cos \Theta + 0.2 \sin \Theta + 0.4 \cos 2\Theta + 0.8 \sin 2\Theta$$

$$\phi_2 = -12.9 \cos \Theta - 1.0 \sin \Theta + 1.8 \cos 2\Theta + 1.4 \sin 2\Theta \quad ,$$

where $\odot$ is the longitude of the sun measured from January 0.0 of the year. These expressions are in dimensionless parameters in parts per 10^8.

We should like, at this point, to compare this with the observed excitation function. To do this, we must evaluate the excitation function from other sources. These sources are classified into five groups: 1) the body tides, 2) the ground water, 3) the response of the oceans to variations other than those due to pressure, 4) the winds, and 5) the currents. The body tides are quite well established and have an effect of the order of 0.05 arcsec, which can be ignored. The winds have been disposed of as essentially geostrophic, and the ocean currents seem to be too small (Munk and MacDonald, 1960, p. 127). The ground water and oceans are estimated by Munk and MacDonald (Chapter 9). Table 4 (from Munk and MacDonald) summarizes these results, with the Munk and Hassen (1961) results included in parentheses.

In compiling the table, Munk and MacDonald executed two calculations. Where possible, they estimated the total effect, as well as each component. As these two estimates did not agree, they argued that the estimate of the total contribution was more reliable. This conclusion was based on the fact that the results agreed better with the astronomical data.

On the other hand, Munk and Hassen (1961) ignored the contribution of the oceans and got a total excitation function of

$$\phi_1 = -2.5 \cos \odot + 0.1 \sin \odot$$

$$\phi_2 = -6.5 \cos \odot + 4.9 \sin \odot \quad .$$

Using the Munk and Hassen values together with the other data from Munk and MacDonald, we now get

$$\phi_1 = -12.1 \cos \odot - 5.4 \sin \odot$$

$$\phi_2 = -11.9 \cos \odot - 3.4 \sin \odot \quad .$$

Table 4. The modified excitation function Ψ_i in parts per 10^8, and the mean water load q_E in g cm^{-2} (from Munk and MacDonald, 1960, p. 130)

	Annual							
	Ψ_1		Ψ_2		Ψ_3		q_E	
Astronomical Observations								
Jeffreys (1962)	-0.5 cos ⊙	-0.5 sin ⊙	-15.5 cos ⊙	+6.8 sin ⊙				
Walker and Young (1957)	-4.8	6.3	- 7.3	15.0				
Bureau Internationale d'Heure					-0.45 cos ⊙	-0.36 sin ⊙		
Smith and Tucker (1953)					-0.33	-0.16		
Geophysical Observations								
1. Tides	order (0.05)		order (0.01)		-0.03	-0.00		
2. Air mass	-1.7 (-1.8)	-0.9 (+0.20)	-16.3 (-12.9)	-1.6 (-1.0)	0.02 (-0.0028)	0.01 (+0.0035)		
2.1 pressure correction	-0.1	0.6	9.1	5.2	-0.01	-0.00		
2.2 water vapor	-0.8	0.4	1.5	0.2	-0.00	-0.00	-0.17 cos ⊙	-0.08 sin ⊙
2.3 isostatic oceans	5.1	1.6	5.1	0.6	0.01	-0.01		
3. Ground water	-0.7	-0.1	6.4	5.9	0.02	0.03	0.19	0.77
3.1 snow	-6.1	-5.8	- 3.3	-3.1	0.02	0.02	0.44	0.41
3.2 vegetation	0.2	0.1	0.3	0.1	0.00	0.00	-0.03	-0.01
4. Oceans	-9.6	-5.5	- 4.4	-8.3	0.01	0.02		
4.1 isostatic (2.3)	(5.1)	(1.6)	(5.1)	(0.6)	(0.01)	(-0.01)		
4.2 nonisostatic	-9.8	-5.4	- 4.6	-8.2	0.01	0.02	-0.14	-0.66
4.3 balance	0.2	-0.1	0.2	-0.1	-0.00	0.00	0.12	-0.03
5. Winds					-0.31	-0.20		
6. Currents					order (0.02)			
1 + 2 + 3 + 4 + 5 + 6	-12.0 (-12.1)	-6.5 (5.4)	-14.3 (-11.9)	-4.0 (-3.4)	-0.29	-0.14	0.00	0.00
	Semiannual							
Astronomical Observations								
Walker and Young (1957)	-0.5 cos 2⊙	-2.9 sin 2⊙	4.4 cos 2⊙	+1.0 sin 2⊙				
Bureau Internationale d'Heure					0.28 cos 2⊙	+0.24 sin 2⊙		
Smith and Tucker (1953)					0.41	-0.06		
Geophysical Observations								
1. Tides	order (0.3)		order (0.1)		0.16	-0.06		
2. Ground water	0.7	-0.2	- 0.4	-0.9	-0.00	-0.01		
3. Winds					0.03	0.01		

Of course, there is not extremely good agreement between the observed excitation by Jeffreys and that by Walker and Young (1955, 1957). Jeffreys' results are in better agreement with other investigations that were based on the independent Greenwich-Washington observations (Munk and MacDonald, 1960, p. 129). Munk and MacDonald admit that item 4.2 from the table is the most unreliable, but the agreement of their solution does not seem very much better without it, being

$$\phi_1 = -2.3 \cos \odot + 0.00 \sin \odot$$

$$\phi_2 = -7.3 \cos \odot - 4.9 \sin \odot \quad .$$

On the other hand, Munk and Hassen feel that the item of greatest uncertainty is ground water (item 3). That there is order-of-magnitude agreement is gratifying; that there is good agreement between the observed excitation function and the geophysically computed excitation function is dubious.

In the next few years the distribution of water in the oceans should be observed much more accurately, and one part of the uncertainty can be eliminated.

6.2 The Chandler Motion

The Chandler wobble has several areas of interest. The 14-month period is explained by adopting some elastic constants for the earth. This 14-month period is, however, not sharply defined. The 1-year period is much more sharply defined (see Figure 3). We can view this broad spectrum as one of periodic responses to a driving force with a broad range of resonance. We can then examine the atmospheric motion for this excitation. In the language of communications engineering, we can view the earth as acting as a narrow band-pass filter, excited by broad-band noise, producing a peaked response, as in Figure 6. The Chandler wobble can then be examined as a resonant amplification of the side band at 0.85 cpy. Munk and Hassen (1961) tested this hypothesis by measuring the atmospheric excitation at 0.85 cpy.

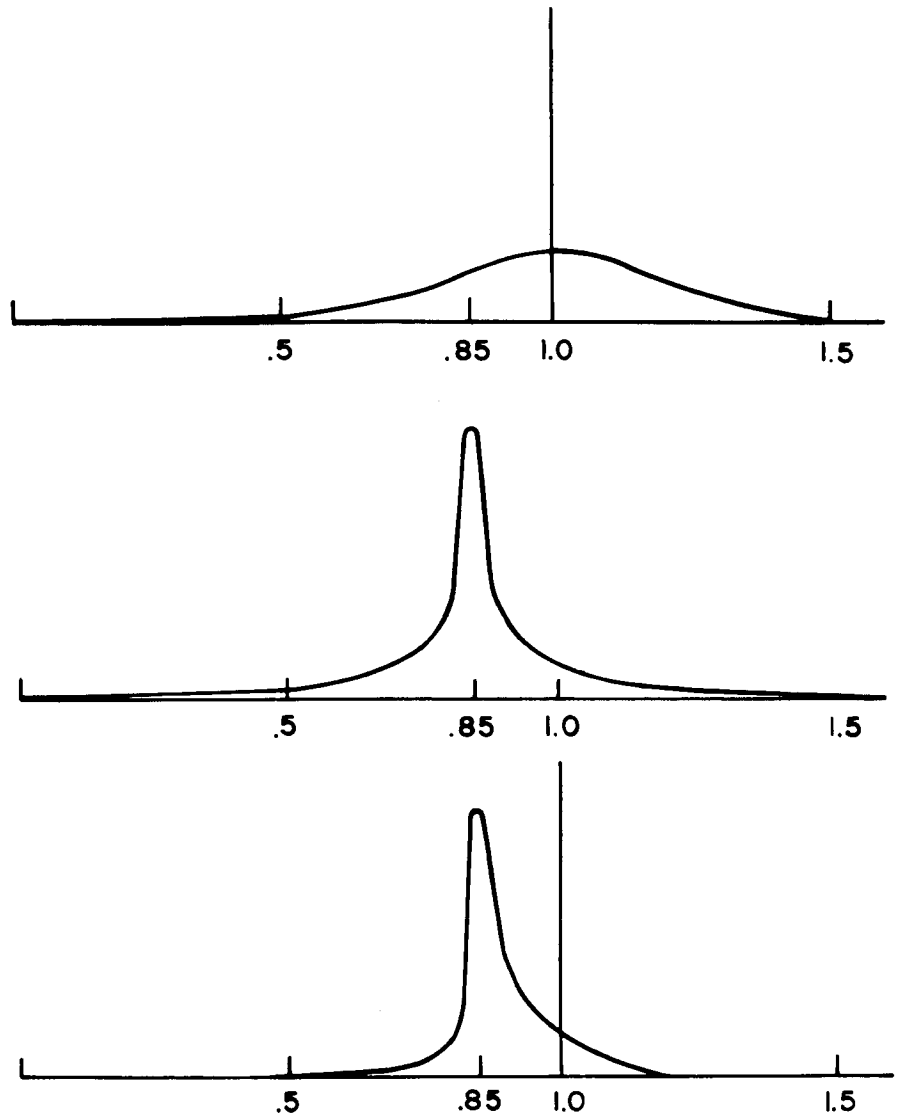


Figure 6. Top: the input spectrum is schematically presented by an annual line plus side bands. Middle: The power transmission of the earth, peaked at the Chandler frequency, 0.85 cpy. Bottom: The power spectrum of the pole of rotation; this curve is the product of the upper two (adapted from Rudnick, 1956). (After Munk and Hassen, 1961.)

Figure 7 compares the observed astronomical excitation with the meteorological excitation. The results are disappointing. The curves bear no resemblance, and in addition, there is an order of magnitude less excitation energy in the atmosphere than in the observed motion. The conclusion by Munk and Hassen (p. 356) is quite negative:

> The smallness of the ratio of computed to observed wobble power implies that the variable distribution of air mass plays a minor role in the excitation of the Chandler wobble. Other variations in the hydrosphere are relatively unimportant to the annual wobble, and can hardly be expected to predominate at 0.85 cpy. Moreover, the low coherence between computed and observed wobble suggests that no other factors in the hydrosphere are of importance.

The motion of the pole is understood in a qualitative way. The annual motion, about 10 feet, is roughly understood, although quantitative calculation of the excitation remains to be brought into agreement with the observed excitation. Questions are still to be resolved in all three major areas, atmospheric motion, ground water, and ocean displacements; none claims to be accurate, each leaving the discrepancy to the others.

An interesting exercise is to evaluate the annual change in pressure, which would explain the polar motion of the 1-year period. This amounts to (Munk and MacDonald, 1960, p. 132)

$$p = \sin\theta \cos\theta [0.14 \cos\lambda \cos(\odot - 45°) + 3.2 \sin\lambda \cos(\odot - 336°)] \quad ,$$

in millibars. This is surely very small compared to the observed variations.

The Chandler motion is yet to be understood. The search for a 0.85-cpy driving force of adequate magnitude is of course predicated on the linear formulation of the equations. Even though the earth may have a 0.85-cpy natural period, the earth must be dissipative. The knowledge of its Q is quite uncertain, and any values between 10 and 1000 would be consistent with available data. In general, 30 is taken as a working number. Therefore, to maintain the 0.85 motion, some energy source must be found. A nonlinear

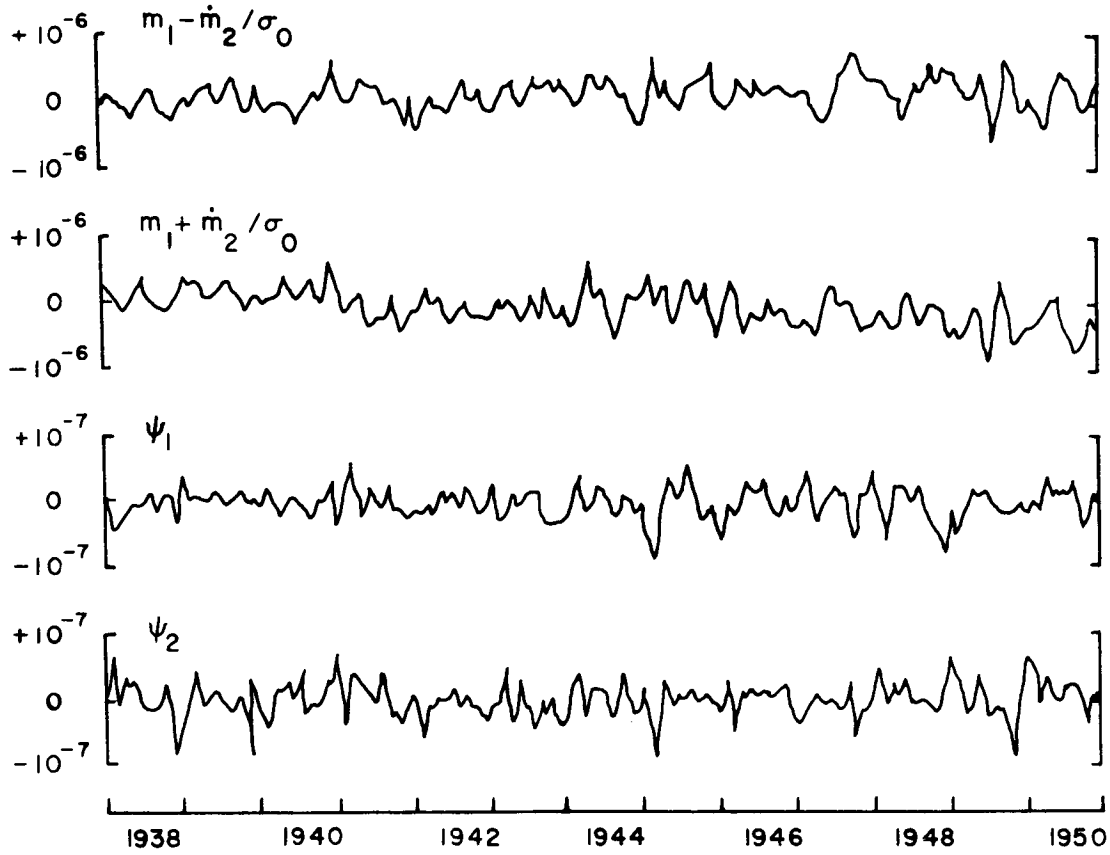


Figure 7. Components of the nonseasonal excitation function inferred from the astronomical observations (upper curves), and derived from meteorological observations (lower curves). The units are in radians; the scale for the meteorological excitation is magnified tenfold relative to astronomically inferred excitation. The position of January means is indicated on the time axis. (After Munk and Hassen, 1961.)

system seems indicated. If indeed this is the situation, many possible sources of energy become available (e. g., the precession). Munk and Hassen (1961) investigated the energy transmitted across the core-mantle boundary, using a linear theory. They ruled this energy out as a source because of the large amount of energy needed for the wobble as compared to that needed for the LOD variations where "one expects the equatorial components to be, if anything, less than the axial component."

In summary, future work must involve more complete observations of the atmosphere, the ocean mass distribution, and the ground water. More detailed geophysical understanding and theoretical developments in the nonlinear mechanics of the earth may also provide some understanding of the earth's wobble.

7. LENGTH OF THE DAY

The variation in the length of the day (for the last 10 years) has been observed by comparing astronomical measurements with modern crystal clocks. Earlier variations (over the last 200 years) were derived from the discrepancy of the lunar longitude based on occultation measurements. The ancient eclipses (2000 years) and Paleozoic corals (1,000,000 years) enable us to deduce the long time-scale variations. We will examine them in this order.

Since the invention of the atomic clock, we have had a daily record of the variation of the LOD. Figure 8, from Markowitz (1966), plots $f = UT2 - A1 + 0.0013\,t + 1$ sec. As we can observe, the curve is reasonably well fit by three parabolas. Of course, there is some question about the irregularities. In July 1959 Danjon claims to have found a discontinuity. Figure 9 shows Danjon's data and those of Ijima and Okazaki (1961). The latter maintain there is no effect. Danjon also claims to have found this effect from Markowitz's data. Figure 10, from Yeh and Shaw (1960), confirms the discontinuity of a change in LOD by 1 msec. Another remarkable set of data, by Shatzman (1966), plots the variation in the LOD against the cosmic-ray index (see Figure 11). The correlation is remarkable. Shatzman proposed that the effect was one of an electromagnetic torque, but his results were negative. The atmosphere could certainly be the vehicle to transmit this effect to the LOD, but this demonstration remains to be done.

The LOD has been studied over the last 250 years by using the lunar occultations to determine the apparent longitude of the moon and by using a lunar theory to compute the actual longitude. The difference between them has been attributed to the variation of the LOD. Figure 12 gives these data. There was certainly a real change in the LOD in 1899.

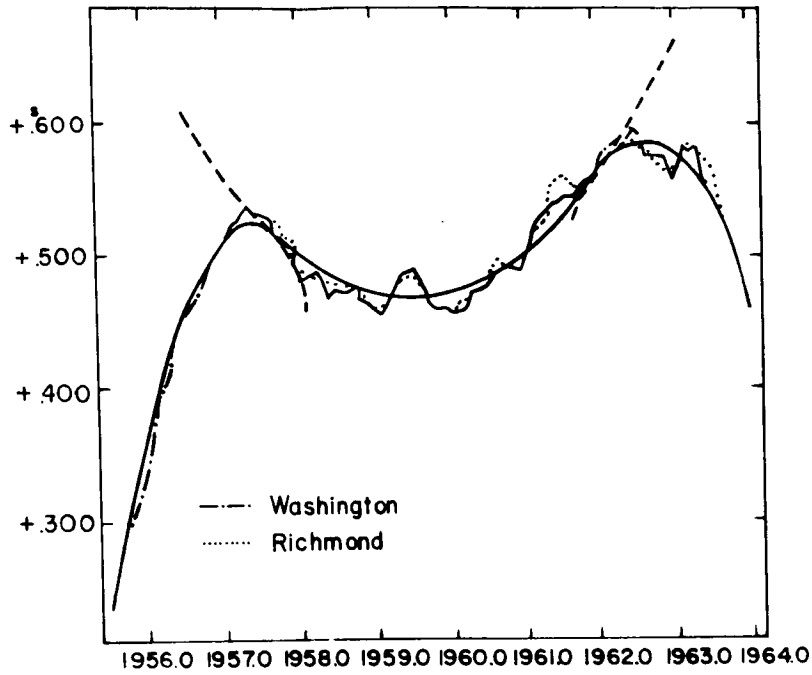


Figure 8. Monthly means of y = UT2 - A1 + 0.0013 t + 1 sec. (After Markowitz, 1966.)

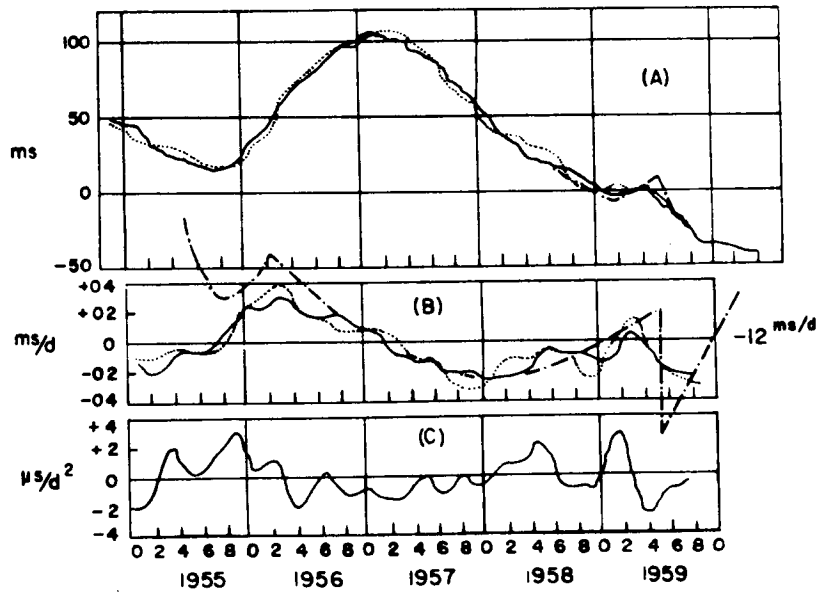


Figure 9. (a) Difference between UT2 and AT (Ijima and Okazaki, 1961). (b) Fluctuation in the rate of rotation of the earth. (c) Fluctuation in the acceleration of rotation of the earth. Solid and dotted curves are for the results with and without, respectively, annual and semiannual seasonal terms. Chain-line curves show the results by Danjon (1962 a, b, c). (After Schatzman, 1966.)

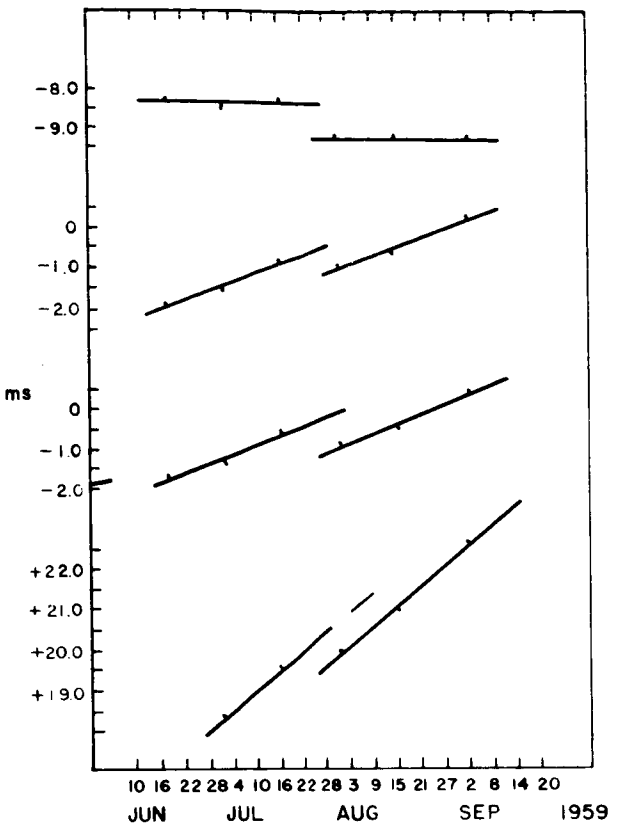


Figure 10. The LOD (Yeh and Shaw, 1960). Four assumptions are made for the continuous effect, but all indicate a discontinuity in July 1959. (After Schatzman, 1966.)

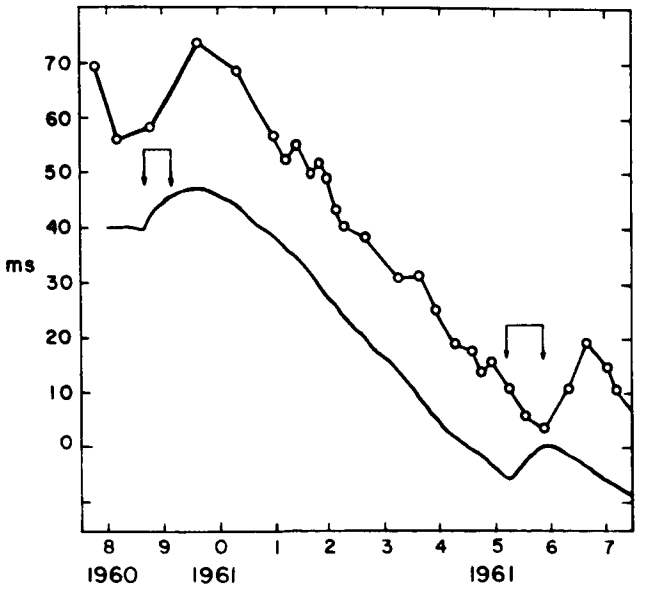


Figure 11. Comparison of f(t) (dotted curve) and a $\int(N_0 - N)dt$ (solid curve). (After Schatzman, 1966.)

The effects described thus far have been essentially irregular in nature. These irregular effects seem to have more intrinsic interest. Going back further in the record, to the ancient eclipses and the Paleozoic corals, the interest was on the most regular effect, the secular deceleration of the earth's motion. Dicky (1966), from a study of ancient eclipses, found that the secular change was

$$\frac{\Delta\dot{\Omega}}{\Omega} = -17.7 \times 10^{-11} \text{ years}^{-1} \quad .$$

The eclipse data may be unreliable, but if true they are certainly very accurate.

The eclipse data allow us to check back several thousand years. A very recent discovery concerning the banding of Devonian fossil corals provides a very exciting look back a million years. Wells (1966) was the first to note that within the annual bands in certain fossils was a smaller banding; he found 400 such bands. He postulated that these were daily variations. Scrutton (1965) has found about 30.5 bands on different corals, possibly due to a monthly variation. Being able to measure the Devonian year in this way allows us to compute the deceleration roughly as $\Delta\dot{\Omega}/\Omega = -27 \times 10^{-10}$ years^{-1}, a value in fair agreement with the eclipse data.

The eclipse data are certainly more accurate today than the fossil data. The relative agreement between the two sets of data is reassuring. The problems of determining the length of the Devonian year by counting fossil bandings, however, are many. The counting of bands is difficult, and laboratory techniques using X-ray defraction are still being developed. There is also a big question about the reaction of the coral to weather patterns. Modern corals exhibit somewhat less than 365 bands. The question of paleoclimatology is of interest now, and the future of this investigation seems very exciting indeed.

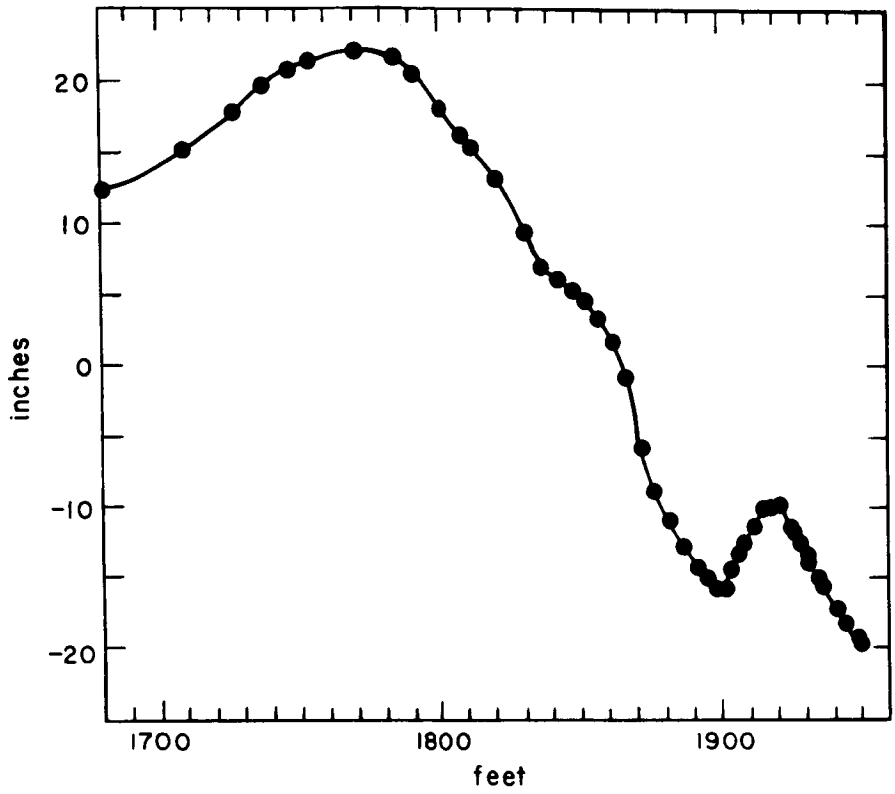


Figure 12. The discrepancy F_ℓ in the longitude of the moon, based on observations of occultations. (After Munk and MacDonald, 1960.)

The short-period effects due to the body tide are reasonably well understood. The slight increase in the tidal effect from summer to winter and the changes in the orientation due to the position of the obliquity produce both an annual and a semiannual change. There are similar effects due to the moon. These tides lead to a change in the moment of inertia, causing the variation in the LOD presented in Table 5.

Table 5. Change in LOD due to body tides

Period	d(LOD) in msec
18.5 years	$0.14 \cos [\beta(\lambda, t)]$
1 year	$0.025 \cos \beta$
0.5 year	$0.15 \cos \beta$
27.55 days	$0.17 \cos \beta$
14.55 days	$0.027 \cos \beta$
13.66 days	$0.33 \cos \beta$
13.63 days	$0.14 \cos \beta$

The secular deceleration of the earth is caused by many effects. The most complete investigation of these effects (Dicky, 1966) is summarized here (see Table 6). The departure from the computed value of $\Delta\dot{\Omega}/\Omega$ is -6.8×10^{-11}. Of course, many of the effects are not well understood, and much work remains to be done in this area.

Table 6. Effects causing the secular deceleration of the earth

Effect	$(\Delta\dot{\Omega}/\Omega)$ year^{-1}
Variations in sea level	-0.1×10^{-11}
Lunar tide	21.0×10^{-11}
Solar tide	6.2×10^{-11}
Atmospheric tide	-1.6×10^{-11}
Short-period core effect	1.2×10^{-11}
Long-period core effect	1.2×10^{-11}
Angular momentum from solar wind	$<10^{-5}$
Angular momentum from atmospheres	$<10^{-12}$
and oceans	$<10^{-13}$
Continental unrest	$< 7 \times 10^{-13}$
Thermal expansion	$< 3 \times 10^{-12}$
Growth of core	$<3.1 \times 10^{-12}$

8. ANOTHER EFFECT

We are all familiar with the idea that the action of the moon and the sun on the oceans causes tides: the resulting motion causes friction, which in turn dissipates energy. This causes the earth's rotation rate to decrease and the moon's distance to increase. Dicky (1966) (see Table 6, items 2 and 3) estimates the amount of deceleration due to these effects. The agreement of these calculations with the estimated age of the earth-moon system is an argument given for the validity of the model.

Jeffreys (1962) and Heiskanen (1921) estimate the amount of energy dissipated in this way. If we apply some averaging, the results are -3.3 and -5.5, respectively. We can adopt the average of these values as

$$\frac{dE}{dt} = -4.4 \times 10^{26} \text{ ergs year}^{-1} \quad .$$

That this tide dissipates energy is due to the fact that the ocean lags the earth.

There has been observed an atmospheric tide that leads the earth, having the effect of accelerating the earth. The energy influx of this tide is estimated to be

$$\frac{dE}{dt} = 6.6 \times 10^{25} \text{ ergs year}^{-1} \quad .$$

This rate is critically dependent on the natural resonance of the atmosphere. There is some question of the existence of this resonance. This free period has been measured by the phase and velocity of long air waves. The Krakatoa eruption of 1883 and the Great Siberian Meteor of 1908 caused debris to be

circulated in the atmosphere, and allowed a determination of the free period of 10-1/2 hours. For this resonance effect to contribute to the energy transfer, it would have to have a period of 12 hours. There is some evidence that the data were consistent with such a period. However, any more precise statements must await more theoretical studies in atmospheric oscillations.

The basic idea, first advance by Holmberg (1952), is that the sun's heating acts as an accelerating force. The atmosphere is compressed, heated, and then released, as in a heat engine. When the resonant period is reached, the energy balance between the tidal dissipation and the heat engine is achieved, and the LOD is stable. The balance does not have to be exact at all times, and in fact today, the decelerating torque is about 10 times the accelerating torque. If this conjecture proves correct, then we must reexamine the entire history of the rotation of the earth.

Finally, it is interesting to detail the energy budget of the earth. Many of the topics in the rotation of the earth involve energy quantities, and any satisfactory explanation of the motions must include the proper figures (see Table 7).

Table 7. Energy budget of the earth

Effect	Amount (ergs year^{-1})
Tidal friction	-4.4×10^{26}
Atmospheric acceleration	6.6×10^{25}
Body tides	1.8×10^{26}
Seismic release	$<1 \quad 10^{25}$
Heat flow (measured)	-9.2×10^{27}
Equivalent J_2 energy	$-2 \quad \times 10^{27}$
Solar radiation on earth	$1 \quad \times 10^{28}$

The energy from body tides comes essentially from the working of the material of the earth by the tidal force. The energy liberated is probably

released through the heat flow. The computed energy depends on the phase of the tide with respect to the deforming potential. This can be only poorly estimated from direct observation. The expression for the energy is $7.33 \times 10^{27} \sin 2\phi$ ergs year^{-1}. For large Q, the damping factor, $\tan 2\phi \cong 1/2Q$. A Q of 40 seems reasonable from the study of the Chandler motion, but the damping factor is very poorly determined and can be as small as 10 and as large as 200.

A very important result from satellite geodesy is the departure of the J_2 of the actual earth from the hydrostatic J_2. This means that there are stresses remaining in the oblateness. MacDonald (1966) computes that there are 2×10^{34} ergs stored in the oblateness. If we adopt a time scale for the nonequilibrium bulge of 10^7 years, then we get the annual release of energy given above. Again, this energy is probably dissipated in the form of heat.

9. SUMMARY

The purpose of this seminar was to introduce some new ideas into the program at SAO. I have used it as a device to examine some areas related to the satellite geodesy program already in progress. With the establishment of the SAO Standard Earth as a reference, almost all areas touched on here can be studied and augmented. I have not made a real effort to relate precisely the geodesy results to the rotation of the earth, as most of these relations are already clear. One exciting aspect is that we can reevaluate many of the accepted results, and either improve them or cast them aside.

Clearly, the use of satellite dynamics to study the solid-earth tides and the time-varying aspect of the gravity field are the first topics to come to mind. The use of the gravity field to study stress differences in the earth has already achieved some attention (MacDonald, 1966). In addition, it is possible to study the LOD with satellite observations, and perhaps make an independent determination of UT1. The use of satellite observations in oceanography can produce very precise data about the variations of oceanic distribution and momentum, which are needed to understand the motion of the pole. The period of free nutation would be governed by a pole tide, which would change our ideas about the rigidity of the earth. It seems possible to study the relation between the motion of the mantle and the crust and that of the core by gravitational perturbations.

With combined electronic and optical data we can improve some of the fundamental constants of the earth-moon system. We hope that knowledge of the present situation of the earth's dynamics will enable us to understand more completely the early history of the earth, and to settle some of the more controversial aspects of the continental-drift hypothesis, in addition, possibly, to measuring the present rate of drift.

However, it would seem wise to understand better what we already observe, e.g., the discrepancy between the various determinations of the annual motion of the pole. Apparently, 65 years of observations are not sufficient, and more data are needed.

In summary, it seems that we are on the verge of much new information about the earth and its environment. The fact that the space program is the vehicle for studying the earth is incidental. What is needed are more geophysicists, to study the seemingly endless and exciting problems of geodynamics.

10. BIBLIOGRAPHY

DANJON, A.

1962a. Astronomie fondamentale — Sur la variation continue de la rotation de la terre. Compt. rend., vol. 254, pp. 2479-2482.

1962b. Astronomie fondamentale. La rotation de la terre et le soleil calme. Compt. rend., vol. 254, pp. 3058-3061.

1962c. Résultats déduit des observations à l'astrolabe o.p.l. de Paris et au p.z.t. de neuchâtel juillet 1956 — mars 1960. Notes et informations. Publ. de l'Observatoire de Paris, VIII, no. 7, pp. 1-29.

DICKE, R. H.

1966. The secular acceleration of the earth's rotation and cosmology. In The Earth-Moon System, ed. by B. G. Marsden, and A. G. W. Cameron, Plenum Press, New York, pp. 98-164.

GOLD, T.

1958a. Irregularities in the earth's rotation. I. Sky and Telescope, vol. 17, pp. 216-218.

1958b. Irregularities in the earth's rotation. II. Sky and Telescope, vol. 17, pp. 284-286.

GROVES, G. W.

1957. Day to day variation of sea level. Meteorol. Monographs, vol. 2, pp. 32-45.

HAUBRICH, R., AND MUNK, W.

1959. The pole tide. Journ. Geophys. Res., vol. 64, pp. 2373-2388.

HEISKANEN,

1921. User den Ein Fluss der Gezeiten auf die säkuläre Acceleratime des Mondes. Ann. Acad. Sci., Fennicae, vol. 18A, pp. 1- .

HOLMBERG, E. R. R.

1952. A suggested explanation of the present value of the velocity of rotation of the earth. Monthly Notices Roy. Astron. Soc., Geophys. Suppl., vol. 6, pp. 325-330.

IIJIMA, S., AND OKAZAKI, S.

1961. On the rotation of the earth at the "July 1959 Event." Publ. Astron. Soc. Japan, vol. 13, pp. 125-128.

JEFFREYS, H.

1916. Causes contributory to the annual variation of latitude. Monthly Notices Roy. Astron. Soc., vol. 76, pp. 499-525.

1962. The Earth. Its Origin, History, and Physical Constitution. 4th ed., Cambridge Univ. Press, London, 438 pp.

JEFFREYS, H., AND VICENTE, R. O.

1957a. The theory of nutation and the variation of latitude. Monthly Notices Roy. Astron. Soc., vol. 117, pp. 142-161.

1957b. The theory of nutation and the variation of latitude: The Roche model core. Monthly Notices Roy. Astron. Soc., vol. 117, pp. 162-173.

JUNG, K.

1956. Figur der Erde. In Hdbk. der Phys., vol. 47, pp. 535-639, ed. by S. Flügge, Springer-Verlag, Berlin.

KOZAI, Y.

1965. Effects of the tidal deformation of the earth on the motion of close earth satellites. Publ. Astron. Soc. Japan, vol. 17, pp. 395-402.

KULIKOV, K. A.

1950. The motion of the poles of the earth and the variation of latitude. Ouspekhi Astr. Nauk., vol. 5, p. 111.

LOVE, A. E. H.

1908. Note on the representation of the earth's surface by means of spherical harmonics of the first three degrees. Proc. Roy. Soc. A, vol. 80, pp. 553-556.

1909. The yielding of the earth to disturbing forces. Proc. Roy. Soc. A, vol. 82, pp. 73-88.

1911. Some Problems of Geodynamics. Cambridge, Univ. Press, Cambridge, England, 180 pp.

1944. A Treatise on the Mathematical Theory of Elasticity, 4th ed., Dover Publ. Co., New York, 643 pp.

MACDONALD, G. J. F.

1966. The figure and long-term mechanical properties of the earth. In Advances in Earth Science, ed. by P. M. Hurley, MIT Press, Cambridge, pp. 199-245.

MARKOWITZ, W.

1966. The irregular variation in the speed of rotation of the earth. In The Earth-Moon System, ed. by B. G. Marsden, and A. G. W. Cameron, Plenum Press, New York, pp. 9-11.

MARSDEN, B. G., AND CAMERON, A. G. W.

1966. The Earth-Moon System. Plenum Press, New York, 288 pp.

MELCHIOR, P.

1957. Latitude variation. In Progress in Physics and Chemistry of the Earth, vol. 2, pp. 212-242, ed. by L. H. Ahrens, F. Press, K. Rankama, S. K. Runcorn, Pergamon Press, New York.

1966. The Earth Tides. Pergamon Press, New York, 458 pp.

MINTZ, Y.

1954. The observed zonal circulation of the atmosphere. Bull. Amer. Meteorol. Soc., vol. 35, pp. 208-214.

MINTZ, Y., AND MUNK, W.

1951. The effect of winds and tides on the length of the day. Tellus, vol. 3, pp. 117-121.

1954. The effect of winds and bodily tides on the annual variation in length of day. Monthly Notices Roy. Astron. Soc., Geophys. Suppl., vol. 6, pp. 566-578.

MUNK, W., AND GROVES, G.

1952. The effect of winds and ocean currents on the annual variation in latitude. Journ. Meteorol., vol. 9, pp. 385-396.

MUNK, W., AND HAUBRICH, R.

1958. The annual pole tide. Nature, vol. 182, p. 42.

MUNK, W. H., AND MACDONALD, G. J. F.

1960. The Rotation of the Earth. Univ. of Cambridge Press, Cambridge, England, 323 pp.

MUNK, W. H., AND MILLER, R. L.

1950. Variation in the earth's angular velocity resulting from fluctuations in atmospheric and oceanic circulation. Tellus, vol. 2, pp. 93-101.

MUNK, W. H., AND HASSEN, S. M.

1961. Atmospheric excitation of the earth's wobble. Geophys. Journ., vol. 4, pp. 339-358.

NEWTON, R. R.

1965. An observation of the satellite perturbation produced by the solar tide. Journ. Geophys. Res., vol. 70, pp. 5983-5989.

ROUTH, E. J.

1955. Dynamics of a System of Rigid Bodies. Part II. Dover Publ. Co., New York.

RUDNICK, P.

1956. The spectrum of the variation in latitude. Trans. Amer. Geophys. Union, vol. 37, pp. 137-142.

RUNCORN, S. K.

1966. Fossil clocks. Trans. Amer. Geophys. Union, vol. 47, pp. 44.

SCHATZMAN, E.

1966. Interplanetary torques. In The Earth-Moon System, ed. by B. G. Marsden, and A. G. W. Cameron, Plenum Press, New York, pp. 12-25.

SCRUTTON, C. T.

1965. Periodicity in devonian coral growth, Palaeontology, vol. 7, pp. 552-558.

SMITH, H. M., AND TUCKER, R. H.

1953. The annual fluctuation in the rate of rotation of the earth. Monthly Notices Roy. Astron. Soc., vol. 113, pp. 251-257.

SPITALER,

1901. Die periodischen luft massenverschiebungen und ihr einfluss auf die Langenanderungen der Erdachse (Breitenschwankungen). Petermanns Mitteilungen Erganzungsband, vol. 29, p. 137.

1897.

Denkschr. Akad. Wiss. Wien, vol. 64, pp. 633-642.

TOMASCHEK, R.

1957. Tides of the solid earth. Hdbk. der Phys., vol. 48/2, pp. 775-845, ed. by S. Flugge, Springer-Verlag, Berlin.

WALKER, A. M., AND YOUNG, A.

1955. The analysis of the observations of the variation of latitude. Monthly Notices Roy. Astron. Soc., vol. 115, pp. 443-459.

1957. Further results on the analysis of the variation of latitude. Monthly Notices Roy. Astron. Soc., vol. 117, pp. 119-141.

WELLS, J. W.

1966. Palentological evidence of the rate of the earth's rotation. In The Earth-Moon System, ed. by B. G. Marsden and A. G. W. Cameron, Plenum Press, New York, pp. 70-81.

WOOLARD, E. W.

1953. Theory of the rotation of the earth around its center of mass. Astronomical Papers of American Ephemeris Nautical Almanac, Gov't. Printing Office, Washington, D.C.

INTRODUCTION TO THE THEORY OF THE EARTH'S MOTION ABOUT ITS CENTER OF MASS

Giuseppe Colombo[1]

1. LIOUVILLE EQUATION

The motion of the earth about its center of mass is a very complicated one, as Veis (these proceedings) has already pointed out. I will add only that studies of this motion have become more and more complex with the increasing accuracy of measurements, and in order to explain this motion, models will need to be increasingly sophisticated.

Our first problem is to define what we intend to consider as the external closed boundary of the earth: The sea and solid earth surface? An altitude of 200 km above sea level, where the density is nearly constant in time and position? The boundary (not yet sufficiently defined) of the magnetosphere?

Let us assume that the boundary is 200 km above sea level. Although the center of mass G is a well-defined geometrical point, its position is not fixed with respect to any observatory on the surface of the earth because the earth is not a rigid body. Nevertheless, the equation

$$\frac{d\vec{S}_G}{dt} = \vec{M}_G^{(e)} \tag{1}$$

holds, where $\vec{S}_G$ is the total angular momentum vector with respect to G, and $\vec{M}_G^{(e)}$ is the torque due to all external forces. The value $\vec{S}_G$ is difficult to

[1] Celestial Mechanician, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts; regularly at University of Padua, Padua, Italy.

determine, since the relative motions of the different constituents of the earth (core, mantle, ocean, atmosphere) have to be considered. Since we are not dealing with a rigid body, it is impossible to define the axis of instantaneous rotation of the earth. It is generally assumed that, apart from the motion of the ocean and the atmosphere, the other relative motions are of low frequency and small amplitude. Following Jeffreys (1959) we may define the instantaneous angular velocity vector of the earth as $\vec{\omega}$, which minimizes the following expression:

$$\int_{\phi} (\vec{\omega} \times \vec{\ell} - \vec{v})^2 \, \mu \, d\phi \quad , \tag{2}$$

where $\vec{\ell}$ is the vector from G to the element $d\phi$, μ is the density of $d\phi$, and $\vec{v}$ is the velocity of $d\phi$ with respect to G. We could call $\vec{\omega}$ the mean instantaneous angular velocity.

There is no problem in defining an exact reference system. In fact, we can choose a reference system based on the location of a particular observatory on the earth (geographical system) or one that uses fixed stars for the orientation of the axis; or else we can choose the instaneous center of mass of the earth and the instantaneous principal axis of inertia of the earth (inertial reference system). However, there are nearly insuperable difficulties when we have to relate the different systems to each other, particularly the local geographical system to the inertial reference system.

Suppose we assume reference systems based on an observatory O and the fixed stars. We can write the equation

$$\frac{d}{dt} \vec{S}_O + \vec{v}_O \times M\vec{v}_G = \vec{M}_O^{(e)} \quad , \tag{3}$$

where $\vec{S}_O$ is the angular momentum of the earth with respect to O, $\vec{v}_O$ and $\vec{v}_G$ are the velocities of O and G with respect to a Newtonian reference system, and M is the mass of the earth.

Concerning $\vec{M}_O^{(e)}$ and $\vec{M}_G^{(e)}$, we have to choose between the "torque approach" and the "momentum approach."

If in the system of the earth we include the ocean and the atmosphere (up to 200 km above sea level), there is a further complication in the computation of $\vec{S}_G$ or $\vec{S}_O$ and of $\vec{v}_O$ and $\vec{v}_G$ if we use equation (3), because the motions of the ocean and the atmosphere have significant components. In addition, there is another complication, arising from our poor knowledge of the core's motion.

The only important external torque is, in this case, the gravitational torque, because the effect of the other external torque, of electromagnetic origin, or of the interaction with interplanetary matter, is evaluated to be negligible in the relatively short time scale (10^2 to 10^3 years) we are considering here.

If we confine the earth to the solid-crust boundary (torque approach), other and perhaps more severe difficulties arise from contact forces experienced by the ocean and the atmosphere on the solid earth.

Since we have to begin a theory somewhere, let us start with a reference system Σ^* centered on the center of mass G. The x_3 axis should be reasonably well aligned with the axis of the figure or the axis of the maximum moment of inertia, which fortunately is close to the axis of instantaneous rotation, as defined above. The x_1 axis lies in the plane through x_3 that contains the Greenwich Observatory.

Since it is reasonable to think that the x_3 axis is quite far from Greenwich, the reference system is, in this way, defined. The local geographical systems and the principal axis of inertia are all moving with respect to this conventional reference system, which also is moving.

We proceed now to give an explicit form to S_G. Let us call $\vec{i}$, $\vec{j}$, $\vec{k}$ the unit vector along the axes of Σ^*; ω_1, ω_2, and ω_3 the components of the angular

velocity $\vec{\omega}$ of Σ^* with respect to a Newtonian reference system; A, B, C, D, E, and F the coefficients of the inertia tensor of the earth; and h_1, h_2, and h_3 the components of the angular momentum due to the velocity of the earth's oceans, atmosphere, and core, and even the earth's mantle, with respect to Σ^*. Then we have

$$S_G = (A\omega_1 - E\omega_3 - F\omega_2)\vec{i} + (B\omega_2 - F\omega_1 - D\omega_3)\vec{j}$$

$$+ (C\omega_3 - D\omega_2 - E\omega_1)\vec{k} + h_1\vec{i} + h_2\vec{j} + h_3\vec{k} \quad . \tag{4}$$

Here A, B, ..., ω_i, and h_i are functions of time; $\vec{i}$, $\vec{j}$, $\vec{k}$ are also moving.

We note that the following assumptions are generally accepted on the basis of observations and depend on the choice of the reference system. All the following functions of t are first-order infinitesimals, where A_0 and C_0 are constant values:

$$\frac{A(t) - A_0}{A_0} \quad , \quad \frac{B(t) - B_0}{B_0} \quad , \quad \frac{C(t) - C_0}{C_0} \quad , \quad \frac{\omega_1}{\omega_3} \quad ,$$

$$\frac{\omega_2}{\omega_3} \quad , \quad \frac{h_i}{C_0\,\omega_3} \quad . \tag{5}$$

Since in equation (1) the first derivative d/dt should be made with respect to a Newtonian reference system, the equation now has to be written in the form

$$\frac{d'\vec{S}_G}{dt} + \vec{\omega} \times \vec{S}_G = \vec{M}_G^{(e)} \quad , \tag{6}$$

where d′ denotes that differentiation is made with respect to the rotating reference system. With the normal notation and neglecting the second-order quantities, we have the following system:

$$\begin{cases} A_0\dot{\omega}_1 + (C_0 - A_0)\omega_2\omega_3 - \dot{E}\omega_3 + D\omega_3^2 = -\dot{h}_1 + \omega_3 h_2 + M_1 \quad , \\ A_0\dot{\omega}_2 + (A_0 - C_0)\omega_1\omega_3 - \dot{D}\omega_3 - E\omega_3^2 = -\dot{h}_2 - \omega_3 h_1 + M_2 \quad , \\ \dot{C}\omega_3 + C\dot{\omega}_3 = M_3 \quad . \end{cases} \tag{7}$$

We will start from this equation, the Liouville equation, in order to study the motion of ω with respect to Σ^*.

2. THE CHANDLER WOBBLE

We will consider here only short-periodic variations of $\vec{\omega}$. In an interval of time in which secular variations of $\vec{\omega}$, as well as of A, B, C, . . . , are negligible, we can write, neglecting second-order quantities,

$$\begin{cases} A_0\dot{\omega}_1 + (C_0 - A_0)\omega_2\omega_0 - \dot{E}\omega_0 + D\omega_0^2 = -\dot{h}_1 + \omega_0 h_2 + M_1 \\ A_0\dot{\omega}_2 + (A_0 - C_0)\omega_1\omega_0 - \dot{D}\omega_0 - E\omega_0^2 = -h_2 - \omega_0 h_1 + M_2 \\ \dot{C}_1\omega_0 + C_0\dot{\omega}_3^1 = M_3 \quad , \end{cases} \tag{8}$$

where A_0, C_0, ω_0 are the constant mean values of A, C, ω_3, and C_1/C_0, ω_3^1/ω_0 are first-order infinitesimal functions of time.

If E, D, h_1, h_2, M_1, and M_2 in the relatively short interval of time can be supposed to depend not on the position of Σ^* but only on the time and on $\vec{\omega}$ and can be supposed to be a linear function of the component of $\vec{\omega}$ with possibly time-dependent coefficients, the system (8) is a linear system and the general solution is the combination of a free and a forced term.

Actually, M_1, M_2, and M_3 do depend on the position of Σ^*, as does the gravitational torque. But in a relatively short interval of time, since the

motion of the symmetry axis is slow and is close to x_3, we can consider $\vec{M}_G^{(e)}$ as a function only of time. The same is generally said for h_1 and h_2, which are mostly due to oceanic and atmospheric meteorological and tidal motions, to sun and moon tides on the solid crust, and finally to the problematic motion of the core. That h_1 and h_2 depend on ω_1, ω_2, and ω_3, at least for the contribution of the core's motion, seems quite certain, but the problem of expressing this dependence is so severe that we are forced to neglect it, at least for the moment.

It is certainly true that E and D do depend on ω_1 and ω_2, since the earth is a deformable body and only a uniform rotational motion about an axis fixed in Σ^*, and therefore in space, can lead to a constant configuration of the inertia ellipsoid with respect to Σ^*. This axis must be a principal axis of inertia for the motion to be dynamically possible when the external torques are zero. Moreover, for the rotational motion to be stable the axis must be an axis of maximum moment of inertia, in the case of a symmetrical body, or internal dissipation must be present, or both. When the angular velocity changes, the centrifugal potential also changes. It is assumed that the change in the axis of rotation is slow; that is, the earth has time to reach the static configuration relative to the centrifugal potential at any instant before the centrifugal potential changes significantly.

The centrifugal potential, if second-order terms are neglected, can be written

$$U = \frac{1}{2}\,\omega_0^2\,(x_1^2 + x_2^2) - \omega_0 x_3\,(\omega_1 x_1 + \omega_2 x_2) \quad . \tag{9}$$

The first term leads to a small constant contribution to the earth's flattening. The second term, the variable term, is the only one we shall consider.

The corresponding deformation gives rise to an exterior gravitational potential. The definition of the Love number k is

$$V = - k \frac{R^5}{r^5} \omega_0 x_3 (\omega_1 x_1 + \omega_2 x_2) \quad , \tag{10}$$

where R is the earth's radius and $r^2 = x_1^2 + x_2^2 + x_3^2$.

But the corresponding terms (in $x_1 x_3$ and $x_2 x_3$) in the expansion of a gravitational potential that departs slightly from spherical symmetry are from MacCullagh's formula,

$$\frac{3G}{r^5} (E x_1 x_3 + D x_2 x_3) \quad , \tag{11}$$

G being the gravitational constant.

By comparing (10) with (11), we have

$$E = - \frac{k \omega_0 \omega_1 R^5}{3G} \quad , \quad \text{and} \quad D = - \frac{k \omega_0 \omega_2 R^5}{3G} \quad , \tag{12}$$

D and E being measures of the distortion of the earth in the yx and xz senses, respectively, caused by the change of $\vec{\omega}$ inside the body.

Before we proceed with the Chandler wobble study, an observation should be made concerning the Love number k, which we assume to be the tidal effective Love number. The Love number is in some way a measure of the earth's shield against a centrifugal potential or against any perturbing potential such as the differential gravity field of the sun or the moon. From the earth's tide and from the Chandler wobble (assuming the present theory is valid), we obtain $k \cong 0.29$. This same value is obtained from the free oscillations of the earth as an elastic body. Evaluations by different authors differ by 1 part in 20, indicating a good agreement (which may be misleading). Apart from the significant contrast, which can be explained in several ways

(Munk and MacDonald, 1960, page 27), of the value obtained from the figure of the earth (k = 0.96) with the value k = 0.30, there can be no doubt that the response of the earth to perturbing forces is not proportional to the amplitude and is not independent of the frequency of perturbations, as we assumed in order to write equation (12).

Now let us substitute expression (12) into the system (8). In order to find the free component of ω_1 and ω_2, let us consider the following associative homogeneous linear differential system:

$$\begin{cases} A_0\dot{\omega}_1 + (C_0 - A_0)\omega_2\omega_0 + \dfrac{k R^5}{3G}\omega_0^2\dot{\omega}_1 - \dfrac{k R^5}{3G}\omega_0^3\omega_2 = 0 \\ A_0\dot{\omega}_2 + (A_0 - C_0)\omega_1\omega_0 + \dfrac{k R^5}{3G}\omega_0^2\dot{\omega}_2 + \dfrac{k R^5}{3G}\omega_0^3\omega_1 = 0 \end{cases} \quad . \tag{13}$$

Solving the system, we have

$$\omega_1 = A \cos \Omega(t - \tau) \qquad \text{and} \qquad \omega_2 = Q \sin \Omega(t - \tau) \quad , \tag{14}$$

where Q and τ are arbitrary constants (the amplitude and the phase of the equatorial component of ω) and

$$\Omega = \omega_0 \frac{3G(C_0 - A_0) - k\omega_0^2 R^5}{3G A_0 + k\omega_0^2 R^5} \quad . \tag{15}$$

The component of the constant angular momentum in this free motion is

$$\vec{S}_G = (A_0\omega_1 - E\omega_0)\vec{i} + (A_0\omega_2 - D\omega_0)\vec{j} + C_0\omega_0\vec{k} \quad , \tag{16}$$

and since E and D are proportional to ω_1 and ω_2, the vector $\vec{S}_G$ is coplanar with $\vec{\omega}$ and $\vec{k}$, as can be readily shown.

If we assume $k = 0.29$, $A_0 = 8.089 \times 10^{44}$ (cgs), $(C_0 - A_0)/A_0 = 1/304.8$, and $\omega_0 = 7.2921 \times 10^{-5}$ sec^{-1}, we find a period of the free wobble:

$$T_c = \frac{2\pi}{\Omega} = 440 \text{ sidereal days} \quad . \tag{17}$$

The motion is regular precessional motion with the axis of precession along $\vec{S}_G$, which is constant in fixed space. In Figure 1 the polhode and herpolhode are drawn. The opening of the polhode, which is described by $\vec{\omega}$ in Σ^*, is 0.3 arcsec. The opening of the herpolhode, described by $\vec{\omega}$ in fixed space, is 0.001 arcsec. It is very important at this point to compare the two openings, because we have the opposite situation for the precessional motion caused by the moon's and the sun's gravitational forces. The vector $\vec{\omega}$ almost does not move in Σ^*, but the motion in fixed space is quite large. The polhode opening is much smaller than that of the herpolhode. This is one reason why the Chandler wobble can be reasonably well separated from all forced motion.

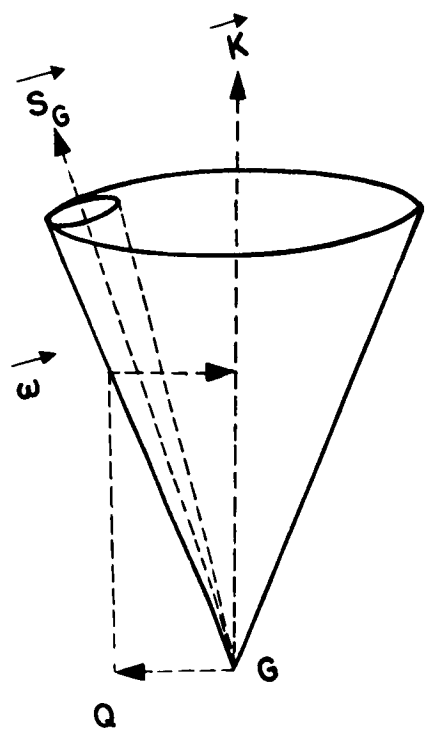


Figure 1. The polhode and the herpolhode.

3. REFERENCES

JEFFREYS, H.

1959. The Earth, its Origin, History, and Physical Constitution. 4th ed., Cambridge Univ. Press, Cambridge, England, 438 pp.

MUNK, W. H., AND MAC DONALD, G. J. F.

1960. The Rotation of the Earth. Cambridge Univ. Press, Cambridge, England, 323 pp.

INTERFACE WITH OCEANOGRAPHY

Walter J. Köhnlein[1]

The tracking of artificial satellites will soon result in position accuracies of about 1 m. In the following we will briefly touch on the impact of satellite tracking on oceanography; however, we will restrict our discussion to physical oceanography, and furthermore, consider only a limited aspect of that — what could be called marine geodesy from space.

Of particular interest in oceanography is the determination of the figure of the coastal lines and its changes with time; the structure of the ocean basins and its stability; and the shape of the ocean surface that is neither an equipotential surface nor any stationary figure, but is continually changing with time owing to the forces acting on it.

Let us start with the coastal lines and the stability of the ocean basin. We do not intend to talk about routine surveying or photogrammetry but will consider movements on a worldwide scale, i. e., the (suspected) radial variation of the earth and the possible measurement of the so-called continental drift. Can we expect to get an answer from position accuracies of 1 m? Because the motions of the continents ("floating" in the denser sima layer) are supposed to reach a few centimeters per year (see Marvin, this report), an intercontinental bridging system performed with lasers and satellites will, at least at the present, have difficulty detecting these small magnitudes. What might lead to measureable effects is a combination of direction and

[1]Geodesist, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts.

time measurements as performed by photo zenith tubes and atomic clocks. Systematic angular displacements of whole sets of stations on different continents would lead to relative changes in latitude and longitude coordinates (see Figure 1) and hence to the detection of a possible continental drift. Markowitz (1948) examined this and other reported drifts, but could not find any such indication. He described it in the following manner:

> Large shifts occurred until the invention of the telegraph, moderate shifts until the invention of radio, and practically none since then.

Perhaps future observations with better distribution and higher accuracy and over longer time intervals can give a more definite answer to this problem.

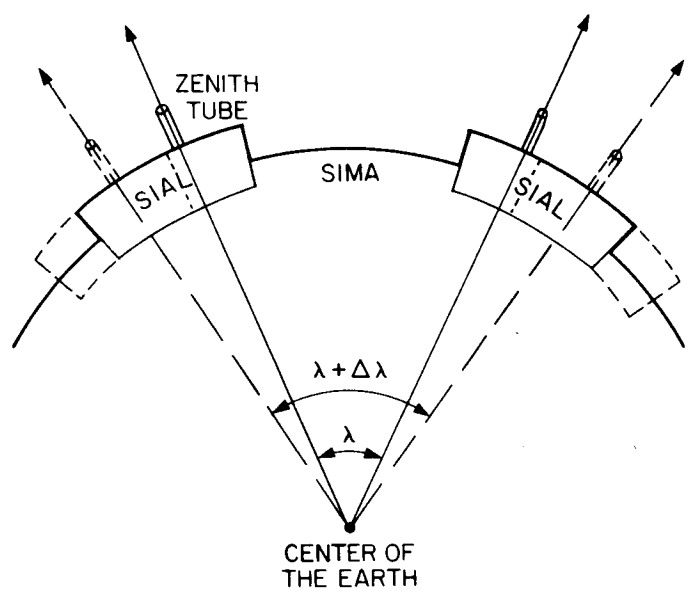


Figure 1. Continental drift as derived from astronomical observations.

Despite this open result, many phenomena point to a continental drift (paleomagnetic comparisons between Europe and North America, geological evidence bearing upon the continental drift, etc.), and hence it is also possible to think of a movement in radial direction; in other words, of an expansion of the earth producing a separation of the continents, as indicated in Figure 2.

This theory is based on a suggestion by Dirac (1938) of a possible decrease in the gravitational constant in the neighborhood of our solar system, so that k in Einstein's field equation

$$R_{ij} - \frac{1}{2} R g_{ij} = -k\, T_{ij} \text{ (standard notation)}$$

is not really a constant, but may be inversely proportional to the age of the universe:

$$-\frac{\dot{k}}{k} = 10^{-9} \text{ to } 10^{-10} \text{ year}^{-1} \quad ,$$

and thus produces a radial expansion of the magnitude of

$$\frac{\dot{a}}{a} \sim 10^{-9} \text{ year}^{-1} \text{ or } \dot{a} \sim 6 \text{ mm year}^{-1} \quad ,$$

according to Jordan (1963), Dicke (1963), and others. With this theory one could immediately explain many geophysical, geological, astronomical, and paleontological phenomena, as for example:

a. The clear-cut separation between continental blocks (sial) and the ocean basins (sima).

b. The worldwide deep-rift system in the oceans and on the continents.

c. The striking geometrical fit of the neighboring shores (including the continental shelves) of Africa and South America.

d. The decrease of the earth's angular velocity (Munk and MacDonald, 1960).

e. The increase of the moon's orbital semimajor axis (Munk and MacDonald, 1960).

f. Paleomagnetic findings, glaciation, volcanic activity, etc.

(A detailed survey is given in Jordan (1963) and in Dicke (1963).) What we are interested in is a possible measurement of a variation of the gravitational "constant" G. In addition to astronomical indications, such as those mentioned in points d. and e., we also could think of evidence from the use of an artificial satellite. Proposals were made to have one satellite accompanied by another, concentric satellite (both being movable against each other by some propulsion sources attached to the outer satellite) in order to avoid the atmospheric drag and the solar-radiation pressure. A tracking network of highest accuracy in direction, distance, and time measurements should be able to attack this problem. We could at least get upper limits of $\dot{G}$ and hence check the possibility of a continental separation in the form of a radial expansion of the earth.

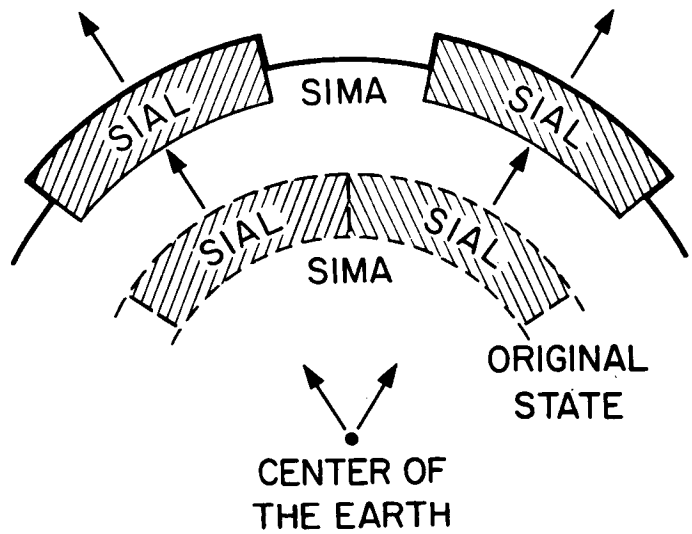


Figure 2. Continental separation by radial expansion of the earth.

Before we proceed, let us review briefly what geodetic information we can obtain from the observation of artificial satellites:

a. Absolute space direction from a Baker-Nunn station to a satellite (Veis, 1963).

b. Absolute space directions between station pairs by simultaneously photographing the satellite against the star background (Veis, 1963).

c. Laser distance measurements between a station and a satellite.

d. Coordinate differences, from the combination of a. or b. with c.

e. The outer gravitational field from orbital analysis (Gaposchkin, 1966).

f. Geocentric coordinates of stations (referred to the mass center of the earth) from orbital analysis (Köhnlein, 1966).

How do the above results relate to oceanographic investigations, assuming we could get an accuracy of 1 m in position? Suppose we have satellite-tracking stations along the coastal lines of the continents; then we can determine their absolute positions in a geocentric coordinate system with their appropriate heights above mean sea level; i.e., we also have the geocentric position of the ocean along the shores. But since the mean sea level is close to the geoid surface (by definition), we can adjust to these points a geocentric ellipsoid and hence obtain the earth's approximate semimajor axis and its flattening for worldwide geodetic systems. Let us go one step further. Because we also know the gravitational field from the motion of satellites, we can compare the shape of an equipotential surface with the actual mean sea level along the coasts, and determine their difference. Because the geoid is defined as the equipotential surface through the mean sea level, we get an approximation of the dimensions of the geoid from what the above-mentioned ellipsoid was — a crude approximation. What we finally need is the actual shape of the whole ocean surface relative to an equipotential surface. Von Arx (1965a) has recently devised new tools to measure the slope of the sea surface relative to the direction of the physical vertical. This certainly could solve the above problem, although the procedure may be too limited

for use on a worldwide scale. Another suggestion is to build an altimeter into a satellite and measure its distance to the sea surface (Moore, 1965). Because the satellite position can be determined relative to tracking stations (simultaneous measurements would even compensate for orbital disturbances), the low-degree slopes of the sea surface (developed into spherical harmonics) could be measured with the accuracy limited only by the tracking network and the precision of the altimeter (see Figure 3). Under very favorable conditions (decimeter accuracy) we might even expect information (Von Arx 1965b) such as tidal phenomena in the open sea, detection of surface waves and tsuamis, correlation of atmospheric changes with the constitution of the ocean surface (atmospheric pressure, temperature, etc.) and rise of sea level across ocean currents. The great advantage of the orbiting altimeter is in repeated measurements of the above phenomena and hence their changes in time. Satellites in particular orbits can serve special purposes; for example, sun-stationary orbits could be especially useful for the detection of lunar tides, and a moon-stationary orbit could be used for the measurement of solar tides.

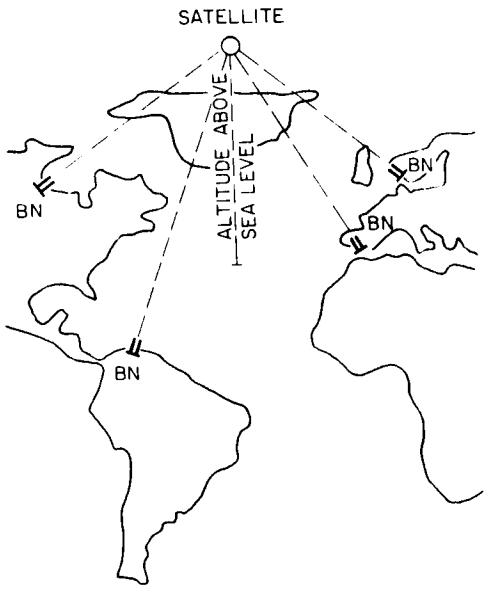


Figure 3. Determination of the low-degree sea-surface slopes by a satellite altimeter.

So far we have mentioned only the vertical component of the sea. Accurate horizontal positions of islands or gyroscopically stabilized platforms on ships are also receiving increasing interest in the space age, as demonstrated by the Apollo ships, designed for guidance of moon and interplanetary vehicles. However, the ocean does not need to be covered with an extensive net of fixed points such as buoys or transponders; in most cases, the accuracy obtained by satellite navigation (Newton, 1966) or a worldwide OMEGA system is wholly sufficient (accurate to between 100 and 400 m). Should the need arise for the highest precision, then the absolute position can be determined once with satellites relative to a net of bench marks; all further positions can be measured with classical means against this fixed point system.

What conclusions can we draw from the above examples? It seems that merely the height component (or geocentric radial distance) of our hypothetical accuracy would already be sufficient for the solution of many problems in (worldwide) marine geodesy. For only a very few limited purposes is a 1 -m accuracy in horizontal position also needed, and this in most cases only insofar as relative positioning is concerned.

REFERENCES

DICKE, R. H.

1963. Experimental relativity. In Relativity, Groups and Topology, ed. by C. De Witt and B. De Witt, Gordon and Breach, Science Publishers, New York, pp. 165-313.

DIRAC, P. A. M.

1938. New basis for cosmology. Proc. Roy Soc. London, vol. 165A, pp. 199-208.

GAPOSCHKIN, E. M.

1966. Tesseral harmonic coefficients and station coordinates from the dynamic method. Smithsonian Astrophys. Obs. Spec. Rep. 200, vol. 2, pp. 105-258.

JORDAN, P.

1963. Four lectures about problems of cosmology. In Cosmological Models, ed. by Centro de Cálculo Cientifico, Instituto Gulbenkian de Ciência, Lisbon, Portugal, pp. 101-136.

KÖHNLEIN, W. J.

1966. Combination of geometric and dynamic results. Smithsonian Astrophys. Obs. Spec. Rep. 200, vol. 2, pp. 259-300.

MARKOWITZ, W.

1945. Redeterminations of latitude and longitude. Trans. Amer. Geophys. Union, vol. 26, pp. 197-199.

MOORE, R. K.

1965. Satellite radar and oceanography, an introduction. In Oceanography from Space, ed. by G. C. Ewing, Woods Hole Oceanographic Institution, Woods Hole, Mass., pp. 355-366.

MUNK, W. H., AND MacDONALD, G. J. F.

1960. The Rotation of the Earth. Cambridge Univ. Press, Cambridge, England, 323 pp.

NEWTON, R. R.

1966. The Navy navigation satellite system. Presented at the VIIth COSPAR International Space Sci. Symp., Vienna, Austria.

VEIS, G.

1963. The determination of absolute directions in space with artificial satellites. Smithsonian Astrophys. Obs. Spec. Rep. No. 133, 24 pp.

VON ARX, W. S.

1965a. Absolute dynamic topography. Soc. Limnol. and Oceanog., vol. 10, Alfred C. Redfield 75th Anniv. Vol. Suppl., pp. R265-R273.

1965b. Orbiting microwave altimeter. Presented at Space Research Summer Study, Working Group in Physical Science, Natl. Acad. Sci., June.

EDITORS' NOTE

The following three papers represent an effort to assess the advisability of preparing computer programs for orbits about the earth with enough generality so that the same programs would be applicable to orbits about the moon or about the sun. Similarly, there arises the question of the extent to which analytical developments for any of these cases are useful in the other cases.

A DIFFERENTIAL ORBIT IMPROVEMENT PROGRAM FOR LUNAR ORBITS[1]

George Veis[2]

The Smithsonian Astrophysical Observatory has developed computer programs, based on a phenomenological approach, to determine orbits of artificial earth satellites from optical-tracking data. The first version of this program — known as Differential Orbit Improvement Program (DOI) — was prepared in the fall of 1958. Since that time the program has undergone many improvements, but the basic principle has not been changed. Detailed description of the program can be found in the literature (Veis, 1960; Veis and Moore, 1960; Gaposchkin, 1966).

The program assumes that the orbit of a satellite is a Keplerian one defined by the Keplerian elements, which, however, are functions of time. These functions are specified with secular (polynomial in time), periodic, and hyperbolic terms. Any element b can be expressed in the form:

$$b = b_0 + \sum P_n(t - t_0)^n + \sum a_i \sin\left[b_i + c_i(t - t_0)\right] + \sum h_j e^{g_j\left[\ell n(k_j - t)\right]} ,$$

where t_0 is the epoch.

[1] Editors' Note: This paper was not presented in exactly this form at the SAO Summer Seminar. However, the contents of the paper were the subject of discussion. Therefore, we felt it desirable to include the paper here. It was presented formally at the International Conference on Selenodesy, Manchester, England, May 28-June 4, 1966.

[2] Geodesist, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts; regularly at National Technical University, Athens, Greece.

Accordingly, the orbit is defined with a set of parameters P_n, a_i, b_i, c_i, h_j, g_j, k_j for each element.

These parameters depend on the perturbing forces (zonal and tesseral harmonics of the earth's gravity field, atmospheric drag, etc.), and, of course, some of them are correlated. However, the program computes the values of the parameters that define each element by a least-squares fitting to the observations that ignores the physical meaning of each parameter.

An initial approximate value for each parameter is given, and the program computes a correction to each parameter by minimizing the sum of the squares of the residuals to the observations. Analytical expressions give the derivatives of each observed quantity with respect to each parameter that describes the elements. The program can handle all forms of geometric observations, i. e., directions expressed in any system, and ranges. Doppler observations (range-rate) are included in the program. These derivatives constitute the elements of the matrix of the equations of condition, which, after normalization, are solved to obtain the corrections to the approximate values. Since the observed quantities are not linear functions of the orbital elements, and since the system is not linear, an iterative procedure is needed. The corrections are added to the approximate values and a new least-squares solution is obtained. The iteration is assumed to converge if the change of the standard deviation on two successive iterations differs by less than 1%.

The exact form and the number of terms for each element that should be used to describe the orbit depend on the satellite and the perturbations affecting its motion, and on the period the orbit is to cover. A rough orbit, however, can be determined using nine parameters, namely a constant eccentricity and inclination, a linear expression in time for the right ascension of the node and the argument of perigee, and a quadratic for the mean anomaly.

If the observations were perfectly distributed, the orbit to be determined would be a mean orbit. Since, however, the distribution of optical observations is far from perfect, modifications had to be made in the program to correct for all short-period effects. The first step was to apply a correction for the short-period perturbations arising from the J_2 term. Furthermore, it was felt that the lunisolar perturbations should be included analytically in the program.

We can say, then, that the program in its present form can apply, as corrections to the observations, all perturbations which can be treated analytically and for the magnitudes of which we have a good knowledge. Actually, both zonal and tesseral harmonics can be treated analytically in the program.

This program is used not only very successfully for the determination of satellite orbits for prediction purposes and all routine tracking operations of the Observatory, but also as the core for orbit-analysis research on the determination of the geopotential and the geocentric coordinates of the tracking stations, and on the study of the upper atmosphere.

The Smithsonian Astrophysical Observatory plans to expand the Differential Orbit Improvement Program to determine lunar orbits. The orbits will be determined in a moon-fixed selenocentric reference system. Observations of direction, range, and range-rate (Doppler) from the earth's surface as well as from the moon's surface will be used. We plan that observations of lunar satellites with respect to the moon will also be employed. An ephemeris for the moon will have to be given, but additional corrections to that ephemeris will be determined in the same way that corrections to the observing stations are determined from the earth satellites.

Since the perturbations arising from the earth (and the sun) will be very important and will considerably distort the orbit, an analytical treatment for those will be applied.

The DOI program was so successful in the determination of the orbits of earth satellites in the early days of the space age because there was no need either to make too many assumptions about the perturbations of the satellite or to have a complete and precise theory. Instead, a mean orbit was fitted to the observations phenomenologically and the analysis was made at a second stage. At that time little was known about the gravity field of the earth (e. g., the odd zonal harmonics) and about how far the upper atmosphere reached.

We expect that we will have to face a similar situation when the first satellites circle the moon. We thus feel that such a phenomenological approach will have many advantages in the determination of selenocentric orbits, at least for the beginning of the lunar satellite era.

REFERENCES

GAPOSCHKIN, E. M.

1964. Orbit determination. Smithsonian Astrophys. Obs. Spec. Rep. No. 200, vol. 1, pp. 77-184.

VEIS, G.

1960. Geodetic uses of artificial satellites. Smithsonian Contr. Astrophys., vol. 3, no. 9, pp. 95-161.

VEIS, G., AND MOORE, C. H.

1960. Smithsonian Astrophysical Observatory differential orbit improvement program. In Astronautical Information, Semiannual Proceedings, JPL, Pasadena, California, pp. 161-184.

THE FORCE FUNCTION ON A LUNAR SATELLITE DUE TO THE OBLATENESS OF THE MOON

Salah E. Hamid[1]

1. INTRODUCTION

The external potential of an oblate body has drawn the attention of almost everyone dealing with artificial-satellite problems. That potential, at a point C at distance r from the center O of the body, can be considered through the analysis of the following function of tesseral harmonics:

$$\frac{1}{r^{\ell+1}} P_\ell^m(\mu) e^{\sqrt{-1}\, m \lambda} \quad , \tag{1}$$

since the imaginary and real parts of that function compromise the main terms of the potential. In expression (1) m and ℓ are nonnegative integers, $\beta = \sin^{-1}(\mu)$, and λ defines the latitude and longitude of the point C referred to the body's equator and a reference meridian.

The expansion, however, of the potential function in terms of β and λ is rather inconvenient when the perturbation effects on a satellite's orbit due to the oblateness of these attracting bodies are being studied. It is usually desirable to express these harmonics in terms of the orbital elements of the satellite. This will enable us to utilize more conveniently the variation-of-elements technique and its well-known related methods in celestial mechanics for calculating the perturbations analytically.

---

[1] Celestial Mechanician, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts.

Such a required process of transformation, in spite of appearing as a direct problem in trigonometry, is easily seen to be intricate when tesseral harmonics of larger values of ℓ and m are considered. Systematic methods for carrying such a transformation have been developed and pursued recently (see, e. g., Kaula, 1961; Izsak, 1964; Garfinkel, 1965). For problems concerning the earth's satellites, usually one transformation is required – that of changing the plane of reference from the earth's equator to the satellite's orbital plane.

For lunar satellites, however, if we take the ecliptic to be the plane of reference for defining the lunar satellite's elements, we need two transformations: the first from the lunar equator as a reference plane to the ecliptic, and the second from the ecliptic to the lunar satellite orbit (Oesterwinter, 1965).

As can be expected, such transformations are more complicated than the one transformation involved in the problems of earth satellites. It would be convenient to have a systematic approach that would give us the lunar effect due to any tesseral harmonics of any degree and order (Jeffreys, 1965).

In the following discussion, we shall outline a systematic method of carrying on these transformations for the lunar satellites. We shall also pursue the required expansions for the evaluation of the potential of an oblate body, using the capability of the digital computer.

2. THE POTENTIAL OF THE MOON'S OBLATENESS

Consider the following potential at point P, due to the oblateness of the moon;

$$R = \mu \sum_{\ell=1}^{\infty} \frac{a^{\ell}}{r^{\ell+1}} \sum_{m=0}^{\ell} (C_{\ell m} \cos m\lambda + S_{\ell m} \sin m\lambda) P_{\ell}^{m}(\sin\beta) \quad , (2)$$

where

μ = $K^2 m_\ell$, K^2 = gravitational constant, m_ℓ = mass of the moon,

a_ℓ = mean equatorial radius of the moon,

r = distance of the point P from the moon's center,

λ, β = selenocentric longitude and latitude of the point P,

$C_{\ell m}, S_{\ell m}$ = constants depending on the mass distribution of the moon.

3. THE FIRST TRANSFORMATION TO THE ECLIPTIC AS PLANE OF REFERENCE

3.1 The Eulerian Angles of the First Transformation

In the above definitions of λ and β, we have the lunar equator and the prime meridian as our planes of reference. This means that our reference set of rectangular axes consists of

OX' = from the center of the moon to the intersection of the prime meridian with the moon's equator,

OZ' = from the center of the moon to the north pole of the moon's equator,

OY' = such that $OX'\ Y'\ Z'$ forms a right-hand system of axes.

The first transformation involved in the evaluation of the potential R is from the lunar equator to the ecliptic. Taking the plane passing through the moon's center and parallel to the ecliptic as our new plane of reference, we can define the following new set of rectangular axes:

OX_0 = from the moon's center to the direction of the vernal equinox,

OZ_0 = from the moon's center to the north pole of the ecliptic,

OY_0 = such that $OX_0\ Y_0\ Z_0$ forms a right-hand system.

Let ON define the direction from the center of the moon to the ascending intersections of the moon's equator with the ecliptic. The rotation from the set of axes OX′ Y′ Z′ to the new set $OX_0\ Y_0\ Z_0$ implies rotation through the following set of Eulerian angles I_0, Ω_0, ω_0 (Jeffreys, 1965), where

$I_0 = -(I_{\text{(}} + \rho)$,

= angle of rotation from OZ′ to OZ_0 measured right-handed about ON ,

$\Omega_0 = 2\pi - (\lambda_{\text{(}} - \Omega_{\text{(}} + \tau - \sigma)$,

= angle X′ON measured right-handed around OZ′

$\omega_0 = 2\pi - (\Omega_{\text{(}} + \pi + \sigma)$

= angle NOX_0 measured right-handed about OZ_0 .

In the above definitions of the Eulerian angle I_0, Ω_0, ω_0, we have

$I_{\text{(}}$ = inclination of the moon's mean equator to the ecliptic,

$\lambda_{\text{(}}$ = mean longitude of the moon,

$\Omega_{\text{(}}$ = ecliptic longitude of the ascending node of the mean lunar equator with the ecliptic,

ρ = physical libration in inclination,

τ = physical libration in longitude,

σ = physical libration in node.

The set of Eulerian angles I_0, Ω_0, ω_0 is related to another set of Eulerian angles I_0, η_0, χ_0 defining the above rotation through the following relations:

I_0 = same as defined above,

$\Omega_0 = \eta_0 + \pi/2$,

$\omega_0 = \chi_0 - \pi/2$.

3.2 The Formula for the First Transformation

With these definitions of the Eulerian angles defining the transformations from the lunar equator to the ecliptic as the plane of reference, we have the following transformation formula, if ϕ', λ' are the latitude and longitude of the direction OP referred to the new set of axes $OX_0 Y_0 Z_0$ (Jeffreys, 1965):

$$P_\ell^m(\sin\beta)\, e^{\sqrt{-1}\, m\lambda} = (\ell - m)!\ (\ell + m)! \sum_{S=-\ell}^{S=+\ell} C_S\, P_\ell^S(\sin\phi')\, e^{\sqrt{-1}\, S\lambda'} \quad .$$

Here

$$C_S = e^{\sqrt{-1}(m\eta_0 + S\chi_0)} \sum_{r=\underline{r}}^{r=\overline{r}} \frac{(-1)^{\ell - m - r}\, (\cos\frac{1}{2} I_0)^{2r+m+S}\, (\sin\frac{1}{2} I_0)^{2(\ell - r) - m - S}}{r!(\ell - m - r)!(\ell - S - r)!\,(m + S + r)!} \quad , \tag{3}$$

where

$\underline{r}$ = lower bound of r = maximum of 0, - m - S, and

$\overline{r}$ = upper bound of r = minimum of ℓ - m, ℓ - S .

Formula (3) can be rewritten in the following form;

$$P_\ell^m(\sin\beta)\, e^{\sqrt{-1}\, m\lambda} = e^{\sqrt{-1}\, m\eta_0} \sum_{S=-\ell}^{S=+\ell} e^{\sqrt{-1}\, S\chi_0}\, P_\ell^S(\sin\phi')\, e^{\sqrt{-1}\, S\lambda'}$$

$$\sum_{r=\underline{r}}^{r=\overline{r}} C'^{\ell\, m}_S(r)\, (\cos\tfrac{1}{2} I_0)^{k_1(r)}\, (\sin\tfrac{1}{2} I_0)^{k_2(r)} \quad , \tag{4}$$

where $C'^{\ell\, m}_S(r)$ means that C' is a function of the parameters ℓ, m, S, and r. In fact the quantities $C'^{\ell\, m}_S(r)$, $k_1(r)$, and $k_2(r)$ can be tabulated once and for all for the various combinations of ℓ, m, S, and r. Notice that the number of

terms involved in the summation over r in expressions (3) and (4) ranges from one term corresponding to the case where $\underline{r} = \overline{r}$, to $\ell - m + 1$ terms corresponding to $\underline{r} = 0$ and $\overline{r} = \ell - m$. Also notice that we always have $k_1(r) + k_2(r) = 2\ell$.

A computer subroutine can be performed to compute $C'^{\ell\,m}_{S}(r)$, $k_1(r)$, and $k_2(r)$ for a given value of ℓ and the corresponding values of m (from $m = 0 \rightarrow \ell$) and S (from $-\ell \rightarrow +\ell$). Table 1 shows part of the output of that subroutine for the values $\ell = 8$, $m = 1$. Notice that the column entitled AK_0 corresponds to value $C'^{\ell\,m}_{S}(r)$.

3.3 Expansion of $\left(\cos\frac{1}{2}I_0\right)^{k_1}\left(\sin\frac{1}{2}I_0\right)^{k_2}$

Now for the lunar satellite, I_0 defined above is given by $I_0 = -(I_\ell + \rho)$. The expression for the lunar libration in inclination consists of the summation of several terms that are functions of the orbital elements of the moon, the sun, and the parameter f that defines the mechanical ellipticity of the moon (Koziel, 1962).

In fact, the general expression for ρ is (Koziel, 1962)

$$\rho = \sum_i \rho_i(f) \cos a_i \quad , \tag{5}$$

where

$$a_1 = \ell_\ell$$
$$a_2 = 2\omega_\ell$$
$$a_3 = \ell_\ell + 2\omega_\ell$$
$$a_4 = 2\ell_\ell + 2\omega_\ell$$
$$a_5 = 3\ell_\ell + 2\omega_\ell$$
$$a_6 = -2\ell_\odot - 2\omega_\odot$$
$$a_7 = \ell - 2\ell_\odot + 2\omega_\ell - 2\omega_\odot$$

Table 1. Transformation from lunar equator to ecliptic — first part.

L	M	S	R	AK_0	K1(R)	K2(R)
8	1	-8	7	1.000000000E 00	7	9
8	1	-7	6	-7.000000000E 00	6	10
8	1	-7	7	9.000000000E 00	8	8
8	1	-6	5	2.100000000E 01	5	11
8	1	-6	6	-4.300000000E 01	7	9
8	1	-6	7	3.600000000E 01	9	7
8	1	-5	4	-3.500000000E 01	4	12
8	1	-5	5	1.890000000E 02	6	10
8	1	-5	6	-2.520000000E 02	8	8
8	1	-5	7	8.400000000E 01	10	6
8	1	-4	3	3.500000000E 01	3	13
8	1	-4	4	-3.150000000E 02	5	11
8	1	-4	5	7.560000000E 02	7	9
8	1	-4	6	-5.880000000E 02	9	7
8	1	-4	7	1.260000000E 02	11	5
8	1	-3	2	-2.100000000E 01	2	14
8	1	-3	3	3.150000000E 02	4	12
8	1	-3	4	-1.260000000E 03	6	10
8	1	-3	5	1.764000000E 03	8	8
8	1	-3	6	-1.320000000E 02	10	6
8	1	-3	7	1.260000000E 02	12	4
8	1	-2	1	7.000000000E 00	1	15
8	1	-2	2	-1.890000000E 02	3	13
8	1	-2	3	1.260000000E 03	5	11
8	1	-2	4	-2.940000000E 03	7	9
8	1	-2	5	2.646000000E 03	9	7
8	1	-2	6	-1.320000000E 02	11	5
8	1	-2	7	8.400000000E 01	13	3
8	1	-1	0	-1.000000000E 00	0	16
8	1	-1	1	6.300000000E 01	2	14
8	1	-1	2	-7.560000000E 02	4	12
8	1	-1	3	2.940000000E 03	6	10
8	1	-1	4	-4.410000000E 03	8	8
8	1	-1	5	2.646000000E 03	10	6
8	1	-1	6	-5.880000000E 02	12	4
8	1	-1	7	3.600000000E 01	14	2
8	1	0	0	-9.000000000E 00	1	15
8	1	0	1	2.520000000E 02	3	13
8	1	0	2	-1.764000000E 03	5	11
8	1	0	3	4.410000000E 03	7	9
8	1	0	4	-4.410000000E 03	9	7
8	1	0	5	1.764000000E 03	11	5
8	1	0	6	-2.520000000E 02	13	3
8	1	0	7	9.000000000E 00	15	1
8	1	1	0	-3.600000000E 01	2	14
8	1	1	1	5.880000000E 02	4	12
8	1	1	2	-2.646000000E 03	6	10
8	1	1	3	4.410000000E 03	8	8
8	1	1	4	-2.940000000E 03	10	6
8	1	1	5	7.560000000E 02	12	4
8	1	1	6	-4.200000000E 01	14	2
8	1	1	7	1.000000000E 00	16	0
8	1	2	0	-8.400000000E 01	3	13
8	1	2	1	8.820000000E 02	5	11
8	1	2	2	-2.646000000E 03	7	9
8	1	2	3	2.940000000E 03	9	7
8	1	2	4	-1.260000000E 03	11	5
8	1	2	5	1.890000000E 02	13	3
8	1	2	6	-7.000000000E 00	15	1
8	1	3	0	-1.260000000E 02	4	12

and for f = 0.73 the values of $\rho_i(f)$ for i = 1 → 3 are given by

$$\rho_1 = -106 \text{ arcsec},$$

$$\rho_2 = +35 \text{ arcsec},$$

$$\rho_3 = -11 \text{ arcsec}.$$

The values for i > 3 are less than 5 arcsec. Confining our analysis to an accuracy of 1 m at the moon's distance implies that we are after an absolute accuracy of the order of 2.5×10^{-9}, requiring an accuracy of the order of 2.5×10^{-5} in the above analysis of the expansion of the potential (Oesterwinter, 1965). This means that we can be satisfied with our expressions for ρ given by the first three terms in equation (5).

With that expansion of ρ and remembering that $I_\ell = 2.7 \times 10^{-2}$ rad, we have

$$\left(\cos\tfrac{1}{2} I_0\right)^{k_1}\left(\sin\tfrac{1}{2} I_0\right)^{k_2} = \frac{1}{2^{k_2}} (I_0)^{k_2}\left[1 - \left(\frac{k_1}{8} + \frac{k_2}{24}\right) I_0^2\right]$$

$$= F(I_0, k_1, k_2) \quad . \tag{6}$$

That expression for $\left(\cos\tfrac{1}{2} I_0\right)^{k_1}\left(\sin\tfrac{1}{2} I_0\right)^{k_2}$ leads to the following cases:

$$k_2 = 0 : F(I_0, k_1, k_2) = 1 - \frac{k_1}{8} I_\ell^2 ,$$

$$k_2 = 1 : F(I_0, k_1, k_2) = -\tfrac{1}{2} I_\ell + \left(\frac{k_1}{16} + \frac{1}{48}\right) I_\ell^3 - \tfrac{1}{2}\rho_1 \cos \ell_\ell$$

$$- \tfrac{1}{2}\rho_2 \cos(\ell_\ell + 2\omega_\ell) - \tfrac{1}{2}\rho_3 \cos(2\ell_\ell + 2\omega_\ell),$$

$$k_2 = 2 : F(I_0, k_1, k_2) = \tfrac{1}{4} I_\ell^2 ,$$

$$k_2 = 3 : F(I_0, k_1, k_2) = -\tfrac{1}{8} I_\ell^3 ,$$

$$k_2 \geq 4 : F(I_0, k_1, k_2) = 0 \quad . \tag{7}$$

3.4 The Final Formula for the First Transformation

With these expansions of $\left(\cos \frac{1}{2} I_0\right)^{k_1} \left(\sin \frac{1}{2} I_0\right)^{k_2}$, we can carry out the summation over r given in equation (4) to obtain the following expansion:

$$P_\ell^m(\sin \beta)\, e^{\sqrt{-1}\, m \lambda} = e^{\sqrt{-1}\, m \eta_0} \sum_{S=-\ell}^{S=+\ell} e^{\sqrt{-1}\, S(\chi_0 + \lambda')} P_\ell^S(\sin \phi') \sum_{r=0}^{3} C''^{\ell\, m}_S(r) \cos(a_r) \quad , \tag{8}$$

where now

$$a_0 = 0 \quad ,$$
$$a_1 = \ell_{(} \quad ,$$
$$a_2 = \ell_{(} + 2\omega_{(} \quad ,$$
$$a_3 = 2\ell_{(} + 2\omega_{(} \quad .$$

The quantity $C''^{\ell\, m}_S(r)$, which again is a function of S, ℓ, m, and r, where now r takes the values 0, 1, 2, 3, can be computed once and for all for the various values of ℓ and for the corresponding values of m (= 0 → ℓ) and S (= $-\ell$ → + ℓ).

Table 2 gives part of the output of a subroutine for computing $C''^{\ell\, m}_S(r)$. The column headings FF(1), FF(2), FF(3), and FF(4) correspond to the quantities $C''^{\ell m}_S(r)$ for r = 1, 2, 3, and 4, respectively.

Table 2. Transformation from lunar equator to ecliptic — second part.

L	M	LS	FF(1)	FF(2)	FF(3)	FF(4)
5	1	-5	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	1	-4	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	1	-3	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	1	-2	-4.9208E-05	0.0000E 00	0.0000E 00	0.0000E 00
5	1	-1	2.7337E-03	0.0000E 00	0.0000E 00	0.0000E 00
5	1	0	-8.0783E-02	1.3904E-03	-2.9380E-04	1.6144E-04
5	1	1	9.9471E-01	0.0000E 00	0.0000E 00	0.0000E 00
5	1	2	5.3865E-02	-9.2696E-04	1.9586E-04	-1.0763E-04
5	1	3	1.0935E-03	0.0000E 00	0.0000E 00	0.0000E 00
5	1	4	9.8415E-06	0.0000E 00	0.0000E 00	0.0000E 00
5	1	5	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	2	-5	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	2	-4	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	2	-3	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	2	-2	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	2	-1	-8.6113E-05	0.0000E 00	0.0000E 00	0.0000E 00
5	2	0	3.8273E-03	0.0000E 00	0.0000E 00	0.0000E 00
5	2	1	-9.4265E-02	1.6222E-03	-3.4276E-04	1.8855E-04
5	2	2	9.9526E-01	0.0000E 00	0.0000E 00	0.0000E 00
5	2	3	4.9414E-02	-6.9522E-04	1.4690E-04	-8.0741E-05
5	2	4	5.4675E-04	0.0000E 00	0.0000E 00	0.0000E 00
5	2	5	2.4604E-06	0.0000E 00	0.0000E 00	0.0000E 00
5	3	-5	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	3	-4	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	3	-3	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	3	-2	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	3	-1	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	3	0	-1.3778E-04	0.0000E 00	0.0000E 00	0.0000E 00
5	3	1	5.1030E-03	0.0000E 00	0.0000E 00	0.0000E 00
5	3	2	-1.0777E-01	1.8539E-03	-3.9173E-04	2.1526E-04
5	3	3	9.9617E-01	0.0000E 00	0.0000E 00	0.0000E 00
5	3	4	2.6957E-02	-4.6348E-04	9.7932E-05	-5.3814E-05
5	3	5	1.8225E-04	0.0000E 00	0.0000E 00	0.0000E 00
5	4	-5	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	4	-4	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	4	-3	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	4	-2	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	4	-1	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	4	0	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	4	1	-2.0667E-04	0.0000E 00	0.0000E 00	0.0000E 00
5	4	2	6.5610E-03	0.0000E 00	0.0000E 00	0.0000E 00
5	4	3	-1.2131E-01	2.0857E-03	-4.4069E-04	2.4216E-04
5	4	4	9.9745E-01	0.0000E 00	0.0000E 00	0.0000E 00
5	4	5	1.3489E-02	-2.3174E-04	4.8966E-05	-2.6907E-05
5	5	-5	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	5	-4	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	5	-3	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	5	-2	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	5	-1	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	5	0	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	5	1	0.0000E 00	0.0000E 00	0.0000E 00	0.0000E 00
5	5	2	-2.9525E-04	0.0000E 00	0.0000E 00	0.0000E 00
5	5	3	8.2012E-03	0.0000E 00	0.0000E 00	0.0000E 00
5	5	4	-1.3489E-01	2.3174E-03	-4.8966E-04	2.6907E-04
5	5	5	9.9909E-01	0.0000E 00	0.0000E 00	0.0000E 00

From Table 2, for example, we have

$$P_5^4(\sin\beta)\, e^{\sqrt{-1}\,4\lambda} = e^{\sqrt{-1}\,4\eta_0}\Big[e^{\sqrt{-1}(\chi_0+\lambda')} P_5^1(\sin\phi')(-2.0667\times 10^{-4})$$

$$+ e^{2\sqrt{-1}(\chi_0+\lambda')} P_5^2(\sin\phi')(6.5610\times 10^{-3})$$

$$+ e^{3\sqrt{-1}(\chi_0+\lambda')} P_5^3(\sin\phi')(-1.2131\times 10^{-1}$$

$$+ 2.0857\times 10^{-3}\cos a_1 - 4.4069\times 10^{-4}\cos a_2$$

$$+ 2.4216\times 10^{-4}\cos a_3)$$

$$+ e^{4\sqrt{-1}(\chi_0+\lambda')} P_5^4(\sin\phi')(9.9745\times 10^{-1})$$

$$+ e^{5\sqrt{-1}(\chi_0+\lambda')} P_5^5(\sin\phi')(1.3489\times 10^{-2}$$

$$- 2.3174\times 10^{-4}\cos a_1 + 4.8966\times 10^{-5}\cos a_2$$

$$- 2.6907\times 10^{-5}\cos a_3)\Big] \quad . \qquad (9)$$

This completes the first transformation from the lunar equator to the plane of the ecliptic as the plane of reference.

We still have to transform from the ecliptic to the lunar satellite orbit as the plane of reference.

4. THE SECOND TRANSFORMATION TO THE LUNAR SATELLITE

4.1 The Eulerian Angles of the Second Transformation

The harmonics $P_\ell^S(\sin\phi')\, e^{\sqrt{-1}\, S\, \lambda'}$ appearing in equation (8), giving the required first transformation, are expressed in the selenocentric ecliptic coordinates of the lunar satellite. As mentioned before, it would be convenient to transform our set of coordinates from the set of rectangular axes $OX_0 Y_0 Z_0$ to the set OXYZ, where

OX = direction from the moon's center to the pericenter of the lunar satellite's orbit,

OZ = direction from the moon's center to the pole of the lunar satellite's orbit,

OY = such that the axes OX, OY, and OZ form a right-hand system of axes.

That transformation entails the rotation of the ecliptic axes OX_0, OY_0, and OZ_0 through the set of Eulerian angles I, η, and χ or I, h, and g, where

I = inclination of the lunar satellite orbit to the ecliptic,

h = $\eta + (\pi/2)$ = longitude of the satellite's orbit's ascending node with the ecliptic,

g = $\chi - (\pi/2)$ = angle between the ascending node (with respect to the ecliptic) of the satellite's orbit and its pericenter.

4.2 The Formula for the Second Transformation

With that new reference plane of the lunar satellite and these new sets of axes, the longitude and the latitude of the lunar satellite will simply be f (the true anomaly of the satellite) and 0.

The tesseral harmoncs $P_\ell^S(\sin\phi')\, e^{\sqrt{-1}\, S\lambda'}$ will degenerate to the following series (Jeffreys, 1965):

$$P_\ell^S(\sin\phi')\, e^{\sqrt{-1}\, S\lambda'} = \sum_{j=0}^{\ell} e^{\sqrt{-1}\left[Sh + (\ell - 2j)(f+g)\right]}\, T_j^{\ell S} \quad , \tag{10}$$

where

$$T_j^{\ell S} = \frac{(\ell - S)!\,(\ell + S)!}{2^\ell\, \ell!\, j!\, (\ell - j)!} \sum_{r=\underline{r}}^{r=\overline{r}} (-1)^{\ell - S - r} (\sqrt{-1})^{\ell - S} \binom{2j}{r} \binom{2\ell - 2j}{\ell - S - r}$$

$$\left(\cos\tfrac{1}{2}I\right)^{\ell + S + 2r - 2j} \left(\sin\tfrac{1}{2}I\right)^{\ell - S - 2r + 2j} \quad , \tag{11}$$

and

$$\underline{r} = \text{maximum of } 0,\ -S - \ell + 2j \ ,$$

$$\overline{r} = \text{minimum of } 2j,\ \ell - S \ . \tag{12}$$

Notice that $T_j^{\ell S}$ is either real or imaginary, depending on whether $\ell - S$ is even or odd.

We define two nonnegative integers k_0, k_1 such that $\ell - S = 2k_0 + k_1$. That means that k_1 is either 0 or 1, depending on whether $\ell - S$ is even or odd. With that definition of k_1, the expression for $T_j^{\ell S}$ can be rewritten in the following more convenient form:

$$T_j^{\ell S} = (\sqrt{-1})^{k_1} \sum_{r=\underline{r}}^{r=\overline{r}} T_{rj}^{\ell S} \left(\cos\tfrac{1}{2}I\right)^{k_1} \left(\sin\tfrac{1}{2}I\right)^{k_2} \quad , \tag{13}$$

where now $T_{rj}^{\ell S}$ is real and k_1, k_2 are functions of ℓ, S, j, and r. Table 3 gives a partial output of a computer subroutine for computing $T_{rj}^{\ell S}$, k_1, and k_2 for a given value of ℓ and the corresponding values of $S(= -\ell \rightarrow + \ell)$ and $j(= 0 \rightarrow \ell)$. The quantity $T_{rj}^{\ell S}$ is given in the table in the column labeled AK_0. The parameter k_1 is given in the column headed Index.

4.3 The Exapnsion of $\left(\cos \frac{1}{2} I\right)^{k_1} \left(\sin \frac{1}{2} I\right)^{k_2}$

In the trignometric function

$$\left(\cos \tfrac{1}{2} I\right)^{k_1} \left(\sin \tfrac{1}{2} I\right)^{k_2} \quad , \tag{14}$$

we note that $k_1 + k_2 = 2\ell$ = even integer. That function can be transformed to the general summation series

$$\sum_{r=0}^{\ell} (a_r \sin r I + b_r \cos r I) \quad . \tag{15}$$

Putting $k_2 = a$, we find that the expressions giving that transformation are the following:

for a = odd ,

$$\frac{2^{2\ell-1} \left(\sin \frac{1}{2} I\right)^{a} \left(\cos \frac{1}{2} I\right)^{2\ell-a}}{(-1)^{(a-1)/2}} = \sum_{S=0}^{\ell-1} (-1)^S \sin (\ell - S) I \sum_{j=0}^{S} \left\{ \binom{a}{S-j} \binom{2\ell-a}{j} (-1)^j \right\} \quad , \tag{16}$$

Table 3. Transformation from ecliptic to lunar satellite orbit.

L	S	J	R	AK0	INDEX	K1	K2
4	-4	0	0	4,37500E 00	0	0	8
4	-4	1	2	1,75000E 01	0	2	6
4	-4	2	4	2,62500E 01	0	4	4
4	-4	3	6	1,75000E 01	0	6	2
4	-4	4	8	4,37500E 00	0	8	0
4	-3	0	0	4,37500E 00	1	1	7
4	-3	1	1	-4,37500E 00	1	1	7
4	-3	1	2	1,31250E 01	1	3	5
4	-3	2	3	-1,31250E 01	1	3	5
4	-3	2	4	1,31250E 01	1	5	3
4	-3	3	5	-1,31250E 01	1	5	3
4	-3	3	6	4,37500E 00	1	7	1
4	-3	4	7	-4,37500E 00	1	7	1
4	-2	0	0	-4,37500E 00	0	2	6
4	-2	1	1	-6,25000E-01	0	0	8
4	-2	1	1	7,50000E 00	0	2	6
4	-2	1	2	-9,37500E 00	0	4	4
4	-2	2	2	-5,62500E 00	0	2	6
4	-2	2	3	1,50000E 01	0	4	4
4	-2	2	4	-5,62500E 00	0	6	2
4	-2	3	4	-9,37500E 00	0	4	4
4	-2	3	5	7,50000E 00	0	6	2
4	-2	3	6	-6,25000E-01	0	8	0
4	-2	4	6	-4,37500E 00	0	6	2

while for a = even,

$$\frac{2^{2\ell-1}\left(\sin\frac{I}{2}\right)^{a}\left(\cos\frac{I}{2}\right)^{2\ell-a}}{(-1)^{a/2}} = \sum_{S=0}^{\ell}(-1)^{S}\cos(\ell - S)I \sum_{j=0}^{S}$$

$$\left\{\binom{a}{S-j}\binom{2\ell-a}{j}(-1)^{j}\right\}$$

$$-\frac{1}{2}\sum_{j=0}^{\ell}\binom{a}{\ell-j}\binom{2\ell-a}{j}(-1)^{j} \quad . \qquad (17)$$

In equations (16) and (17) the value of $\binom{j_1}{j_2} = 0$ whenever $j_2 > j_1$. Notice that for odd values of k_2 the coefficients b_r in expression (15) will vanish, while for even values of k_2 the coefficient a_r will vanish.

The transformation from expression (14) to expression (15) can be carried out through a computer subroutine for any values of k_1 and k_2, provided that

$$k_1 \geq 0 \text{ and } k_2 \geq 0 \text{ and } k_1 + k_2 = 2\ell \quad . \qquad (18)$$

Partial output of such a subroutine is given in Table 4. Notice that a_r and b_r in expression (15) are the results of the division of two integers. In fact, we can write

$$a_r = \frac{a_r^1}{c_r} \qquad b_r = \frac{b_r^1}{c_r} \quad , \qquad (19)$$

where a_r^1, b_r^1, and c_r are integers. Table 4 gives these quantities.

Table 4. Transformation of inclination powers to sums.

512*(SINE(I/2)** 0)(COS(I/2)** 10)=				
1*COS(5I) +	10*COS(4I) +	45*COS(3I) +	120*COS(2I) +	210*COS(1I) +
126*COS(0I) +				
512*(SINE(I/2)** 1)(COS(I/2)** 9)=				
1*SIN(5I) +	8*SIN(4I) +	27*SIN(3I) +	48*SIN(2I) +	42*SIN(1I) +
512*(SINE(I/2)** 2)(COS(I/2)** 8)=				
-1*COS(5I) +	+6*COS(4I) +	+13*COS(3I) +	+8*COS(2I) +	14*COS(1I) +
14*COS(0I) +				
512*(SINE(I/2)** 3)(COS(I/2)** 7)=				
-1*SIN(5I) +	+4*SIN(4I) +	-3*SIN(3I) +	8*SIN(2I) +	14*SIN(1I) +
512*(SINE(I/2)** 4)(COS(I/2)** 6)=				
1*COS(5I) +	2*COS(4I) +	-3*COS(3I) +	+8*COS(2I) +	2*COS(1I) +
6*COS(0I) +				
512*(SINE(I/2)** 5)(COS(I/2)** 5)=				
1*SIN(5I) +	-0*SIN(4I) +	-5*SIN(3I) +	-0*SIN(2I) +	10*SIN(1I) +
512*(SINE(I/2)** 6)(COS(I/2)** 4)=				
-1*COS(5I) +	2*COS(4I) +	3*COS(3I) +	+8*COS(2I) +	-2*COS(1I) +
6*COS(0I) +				
512*(SINE(I/2)** 7)(COS(I/2)** 3)=				
-1*SIN(5I) +	4*SIN(4I) +	-3*SIN(3I) +	+8*SIN(2I) +	14*SIN(1I) +
512*(SINE(I/2)** 8)(COS(I/2)** 2)=				
1*COS(5I) +	+6*COS(4I) +	13*COS(3I) +	-8*COS(2I) +	-14*COS(1I) +
14*COS(0I) +				
512*(SINE(I/2)** 9)(COS(I/2)** 1)=				
1*SIN(5I) +	+8*SIN(4I) +	27*SIN(3I) +	+48*SIN(2I) +	42*SIN(1I) +
512*(SINE(I/2)** 10)(COS(I/2)** 0)=				
-1*COS(5I) +	10*COS(4I) +	+45*COS(3I) +	120*COS(2I) +	-210*COS(1I) +
126*COS(0I) +				

4.4 The Final Formula for the Second Transformation

From the above derived expressions for $\left(\cos \frac{1}{2} I\right)^{k_1} \left(\sin \frac{1}{2} I\right)^{k_2}$, expression (13) for $T_j^{\ell S}$ can be written in the following form:

$$T_j^{\ell S} = (\sqrt{-1})^{k_1} \sum_{r=0}^{\ell} \left[t_j^{\ell S}(0, r) \cos r I + t_j^{\ell S}(1, r) \sin r I \right] . \qquad (20)$$

A computer subroutine can compute the coefficients $t_j^{\ell S}(0, r)$ and $t_j^{\ell S}(1, r)$ for a given value of ℓ and the corresponding values of S (from $-\ell \rightarrow +\ell$), j (from $0 \rightarrow \ell$), and r (from $0 \rightarrow \ell$). Substituting expression (20) for $T_j^{\ell S}$, expression (10) for $P_\ell^S(\sin \phi') e^{\sqrt{-1} S \lambda'}$, in expression (8) for $P_\ell^m(\sin \beta) e^{\sqrt{-1} m \lambda}$, we have

$$P_\ell^m(\sin \beta) e^{\sqrt{-1} m \lambda} = e^{\sqrt{-1} m \eta_0} (\sqrt{-1})^{k_1} \sum_{S=-\ell}^{S=+\ell} e^{\sqrt{-1} S(\chi_0 + h)}$$

$$\left[\sum_{r=0}^{3} C''^{\ell m}_S(r) \cos (a_r) \right] \cdot \sum_{j=0}^{\ell} e^{\sqrt{-1}(\ell - 2j)(f+g)}$$

$$\sum_{r=0}^{\ell} \left[t_j^{\ell S}(0, r) \cos r I + t_j^{\ell S}(1, r) \sin r I \right] , \qquad (21)$$

which can be rewritten in the more convenient form

$$P_\ell^m(\sin \beta) e^{\sqrt{-1} m \lambda} = \sum_{S=-\ell}^{S=+\ell} \text{PART 1 } (\ell, m, S) \cdot \text{ PART 2 } (\ell, S) , \qquad (22)$$

where

$$\text{PART 1 } (\ell\ ,\ m,\ S) = e^{\sqrt{-1}\left(m\eta_0 + S(\chi_0 + h)\right)} \sum_{r=0}^{3} C''^{\ell m}_{S}(r)\ \cos a_r$$

$$\text{PART 2 } (\ell\ ,\ S) = (\sqrt{-1})^{k_1} \sum_{j=0}^{\ell} e^{\sqrt{-1}(\ell - 2j)(f+g)} \sum_{r=0}^{\ell}$$

$$\left[t_j^{\ell S}(0, r) \cos r\, I + t_j^{\ell S}(1, r) \sin r\, I \right] . \tag{23}$$

5. EXPANSION OF PART 1 (ℓ , m, S)

From our earlier definitions we have the following expressions for η_0 and χ_0:

$$\eta_0 = \frac{3\pi}{2} - (\lambda_{\ell} - \Omega_{\ell}) - (\tau - \sigma) \ ,$$

$$\chi_0 = \frac{3\pi}{2} - \Omega_{\ell} - \sigma \ . \tag{24}$$

Defining the $\Theta(mS)$ such that

$$\Theta(mS) = m\lambda_{\ell} - (m - S)\Omega_{\ell} - Sh \ , \tag{25}$$

and putting

$$F_r = C''^{\ell m}_{S}(r) \ , \tag{26}$$

we have the following convenient expression for PART 1 (ℓ , m, S):

$$\text{PART 1 } (\ell,\ m,\ S) = e^{\frac{\pi}{2}(3m - S)\sqrt{-1}}\ e^{-\Theta(mS)\sqrt{-1}}\ e^{-\left(m\tau + (S - m)\sigma\right)\sqrt{-1}}$$

$$\cdot \sum_{j=0}^{3} F_j \cos a_j \ . \tag{27}$$

The physical libration in longitude τ and node σ takes the following form (Koziel, 1962):

$$c = \sum_j \tau_j \sin \ell_\tau(j) \quad ,$$

$$\sigma = \frac{1}{I_\ell} \sum_j \sigma_j \sin \ell_\sigma(j) \quad , \tag{28}$$

where $\ell_\tau(j), \ell_\sigma(j)$ are angles depending on the orbital elements of the moon and the earth, and time.

Numerical values of the coefficients τ_j, I_ℓ σ_j, and the expressions for $\ell_\tau(j)$ and $\ell_\sigma(j)$ are given by Koziel (1962). The coefficients τ_j and σ_j decrease as j increases.

Neglecting squares, products, and higher powers of τ_j, σ_j, and putting

$$\frac{\sigma_j}{I_\ell} = \sigma'_j \quad , \tag{29}$$

we have the following expression for PART 1 (ℓ, m, S):

$$e^{-\frac{\pi}{2}(3m - S)\sqrt{-1}} \quad \text{PART 1 } (\ell, m, S)$$

$$= \left\{ \cos[\Theta(mS)] - \sqrt{-1} \sin[\Theta(mS)] \right\} \left\{ 1 - \sqrt{-1} \left[m \sum_j \tau_j \sin \ell_\tau(j) \right.\right.$$

$$\left.\left. + (S - m) \sum_j \sigma'_j \sin \ell_\sigma(j) \right] \right\} \left(\sum_{r=0}^{3} F_r \cos a_r \right)$$

$$= \left(\sum_{r=0}^{3} F_r \cos a_r \right) \quad .$$

$$\times \left[\left(\cos[\Theta(mS)] - \frac{m}{2}\sum_j \tau_j \left\{\cos[\Theta(mS) - \ell_\tau(j)] - \cos[\Theta(mS) + \ell_\tau(j)]\right\}\right.\right.$$

$$\left. - \frac{S-m}{2}\sum_j \sigma'_j \left\{\cos[\Theta(mS) - \ell_\sigma(j)] - \cos[\Theta(mS) + \ell_\tau(j)]\right\}\right)$$

$$-\sqrt{-1}\left(\sin[\Theta(mS)] - \frac{m}{2}\sum_j \tau_j \left\{\sin[\Theta(mS) - \ell_\tau(j)] - \sin[\Theta(mS) + \ell_\tau(j)]\right\}\right.$$

$$\left.\left. - \frac{S-m}{2}\sum_j \sigma'_j \left\{\sin[\Theta(mS) - \ell_\sigma(j)] - \sin[\Theta(mS) + \ell_\sigma(j)]\right\}\right)\right] . \quad (30)$$

Neglecting $\tau_j F_j$ and $\sigma'_j F_j$ for $j \geq 1$, we have

$$e^{-\frac{\pi}{2}(3m - S)\sqrt{-1}} \quad \text{PART 1 } (\ell, m, S)$$

$$= \left(\frac{1}{2}\sum_{j=0}^{3} F_j \left\{\cos[\Theta(mS) - a_j] + \cos[\Theta(mS) + a_j]\right\}\right.$$

$$- \frac{m}{2}\sum_{j=1} F_0 \tau_j \left\{\cos[\Theta(mS) - \ell_\tau(j)] - \cos[\Theta(mS) + \ell_\tau(j)]\right\}$$

$$\left. - \frac{S-m}{2}\sum_{j=1} F_0 \sigma'_j \left\{\cos[\Theta(mS) - \ell_\sigma(j)] - \cos[\Theta(mS) + \ell_\sigma(j)]\right\}\right)$$

$$-\sqrt{-1}\left(\frac{1}{2}\sum_{j=0}^{3} F_j \left\{\sin[\Theta(mS) - a_j] + \sin[\Theta(mS) + a_j]\right\}\right.$$

$$- \frac{m}{2}\sum_{j=1} F_0 \tau_j \left\{\sin[\Theta(mS) - \ell_\tau(j)] - \sin[\Theta(mS) + \ell_\tau(j)]\right\}$$

$$\left. - \frac{S-m}{2}\sum_{j=1} F_0 \sigma'_j \left\{\sin[\Theta(mS) - \ell_\sigma(j)] - \sin[\Theta(mS) + \ell_\sigma(j)]\right\}\right) , \quad (31)$$

which can be written:

$$e^{-\frac{\pi}{2}(3m-S)\sqrt{-1}} \text{ PART 1 } (\ell,\ m,\ S) = \text{PART 01 } (\ell,\ m,\ S) - \sqrt{-1}\ \text{PART 02 } (\ell,\ m,\ S)\ , \quad (32)$$

where

$$\begin{aligned}\text{PART 01 } (\ell, m, S) \\ \text{PART 02 } (\ell, m, S)\end{aligned} = \frac{1}{2}\sum_{j=0}^{3} F_j \left\{ \begin{matrix}\cos \\ \sin\end{matrix} [\Theta(mS) - \alpha_j] + \begin{matrix}\cos \\ \sin\end{matrix} [\Theta(mS) + \alpha_j] \right\}$$

$$- \frac{m}{2} \sum_{j=1} F_0\, \tau_j \left\{ \begin{matrix}\cos \\ \sin\end{matrix} [\Theta(mS) - \ell_\tau(j)] - \begin{matrix}\cos \\ \sin\end{matrix} [\Theta(mS) + \ell_\tau(j)] \right\}$$

$$- \frac{S-m}{2} \sum_{j} F_0\, \sigma'_j \left\{ \begin{matrix}\cos \\ \sin\end{matrix} [\Theta(mS - \ell_\sigma(j)] - \begin{matrix}\cos \\ \sin\end{matrix} [\Theta(mS) + \ell_\sigma(j)] \right\} \ . \quad (33)$$

Examining more closely the above expressions for PART 1 (ℓ, m, S) and remembering the definition of $\Theta(mS)$, α_j, $\ell_\tau(j)$, and $\ell_\sigma(j)$, we can in general write for PART 1 (ℓ, m, S):

$$\text{PART 1 } (\ell, m, S) = e^{+\frac{\pi}{2}(3m-S)\sqrt{-1}} \sum_j \Gamma_j [\cos (a_j + b_j\, d - Sh) - \sqrt{-1} \sin (a_j + b_j\, d - Sh)\ , \quad (34)$$

where

Γ_j is real and is a function of ℓ, m, S besides j ,

$\left.\begin{matrix} a_j \\ b_j \end{matrix}\right\}$ are real and are functions of m, S besides j ,

d = time .

In writing the above expression for PART 1 (ℓ, m, S), we assumed that the angular parameters related to the orbits of the moon and the earth, and the positions of the moon and the earth are all linear functions of time d. The coefficients of these functions can be obtained from the theories of the moon and the earth.

It can be seen that subroutines can be constructed to build up the quantities Γ_j, a_j, b_j as functions of ℓ, m, S, as well as of j and the orbital elements of the moon and the earth and the epoch of our analysis.

6. EXPANSION OF PART 2(ℓ, m, S)

The expression for PART 2 (ℓ, S) given in equation (23) can be rewritten in the following form:

for ℓ = odd,

$$\text{PART 2 } (\ell, S) = (\sqrt{-1})^{k_1}\left[\sum_{k=\ell,\,\ell-2,\,\cdots 1} \cos ku\, F^{\ell S}_{1k}(I) + \sqrt{-1} \sum_{k=\ell,\,\ell-2,\,\cdots 1} \sin ku\, F^{\ell S}_{2k}(I)\right] , \tag{35}$$

where

$$u = f + g$$

$$F^{\ell S}_{1k}(I) = \left[t^{\ell S}_{\frac{\ell-k}{2}}(0r) + t^{\ell S}_{\frac{\ell+k}{2}}(0r)\right]\cos r I + \left[t^{\ell S}_{\frac{\ell-k}{2}}(1r) + t^{\ell S}_{\frac{\ell+k}{2}}(1r)\right]\sin r I ,$$

$$F^{\ell S}_{2k}(I) = \left[t^{\ell S}_{\frac{\ell-k}{2}}(0r) - t^{\ell S}_{\frac{\ell+k}{2}}(0r)\right]\cos r I + \left[t^{\ell S}_{\frac{\ell-k}{2}}(1r) - t^{\ell S}_{\frac{\ell+k}{2}}(1r)\right]\sin r I ; \tag{36}$$

for ℓ = even ,

$$\text{PART } 2(\ell, S) = (\sqrt{-1})^{k_1}\left[\sum_{k=\ell,\ell-2,\cdots,0} \cos ku\, F_{1k}^{\ell S}(I) + \sqrt{-1}\sum_{k=\ell,\ell-2,\cdots,0} \sin ku\, F_{2k}^{\ell S}(I)\right] - (\sqrt{-1})^{k_1}\frac{1}{2}F_{10}^{\ell S}(I) \qquad (37)$$

Equations (34), (36), and (37) give us the final expressions for PART 1 (ℓ, m, S) and PART 2(ℓ, S), which in turn compromise the main expressions, given through equations (22) and (2), of the potential due to the oblateness of the moon.

The above analysis gives an outline for the derivation of the expression for the potential due to the moon's oblateness at a given direction, in terms of the canonical variables of the lunar satellite.

Anyone who has some familiarity with the capability of the digital computer can see at once how these final formulas can be developed on the computer. A computer program can be written to give the different terms of the potential that are due to the moon's oblateness and that arise from the harmonics $P_\ell^m(\mu)\, e^{i\lambda}$ for any value of ℓ, m.

Such a general program is within the direct capability of the computer, and I hope to work on it in the near future.

In our lunar satellite analysis, in addition to the oblateness of the moon we still have to evaluate the general expression for the attraction of the earth, the sun, and the moon (considering the moon as a sphere); i. e., we must obtain the different components of the force function of a lunar satellite moving in the system sun-earth-moon.

Again, all these different expressions can be evaluated up to any order of magnitude through computer programs. The problem of solving for the satellite's orbit by means of these computer-derived expressions for the force function is a natural extension of the above work.

Von Zeipel's method of solution has proved to be a very powerful analytical approach. The systematic processing of that method on the computer is still to be seen. I hope to find such a systematic approach, one that is not confined only to Von Zeipel's but also includes other methods of dealing with perturbations to be tried, tested, and compared with each other.

7. ACKNOWLEDGMENT

I thank Dr. Charles A. Lundquist for suggesting this problem and for drawing my attention to Oesterwinter's work.

8. REFERENCES

GARFINKEL, B.

1965. Tesseral harmonics perturbations of an artificial satellite. Astron. Journ., vol. 70, pp. 784-786.

IZSAK, I. G.

1964. Tesseral harmonics of the geopotential and corrections to station coordinates. Journ. Geophys. Res., vol. 69, pp. 2621-2630.

JEFFREYS, B.

1965. Transformation of tesseral harmonics under rotation. Geophys. Journ., vol. 10, pp. 141-145.

KAULA, W. M.

1961. Analysis of gravitational and geometric aspects of geodetic utilization of satellites. Geophys. Journ., vol. 5, pp. 104-133.

KOZIEL, K.

1962. Libration of the moon. In Physics and Astronomy of the Moon, ed. by Z. Kopal, Academic Press, New York, pp. 27-59.

OESTERWINTER, K.

1965. The Motion of a Lunar Satellite. Ph. D. Thesis, Department of Astronomy, Yale University, New Haven, Conn.

INTERFACE OF SATELLITE TRACKING AND PLANETARY ORBITS

Jean Meffroy[1]

Refinement in satellite tracking leads to a renewal of the old planetary theories. On the advice of the late Professor Dirk Brouwer, I investigated the building of a first-order general planetary theory through Von Zeipel's method. I considered only one disturbing planet. I referred the disturbed planet to the orbital plane of the disturbing planet and I measured the longitudes from the longitude of the ascending node of the disturbed planet. I neglected the powers of eccentricities and mutual inclination higher than the third. I expressed the indirect part of the disturbing function through Newcomb operators and the principal part of the disturbing function through Newcomb operators and Laplace coefficients.

A.

a. The first step in the building of the theory is the elimination of the short-periodic terms. This elimination transforms the system of canonical equations whose Hamiltonian is the indirect part of the disturbing function into a system of canonical equations whose Hamiltonian is identically equal to zero; it transforms the system of canonical equations whose Hamiltonian is the principal part of the disturbing function into a system T of canonical equations whose Hamiltonian is the sum of four secular terms and one long-periodic term.

b. The Computations Division of SAO checked the Laplace coefficients and their derivations with respect to $D = \alpha\, d/d\alpha$ ($\alpha = a_1/a_2$, a_1 = semimajor axis of the disturbed planet, a_2 = semimajor axis of the disturbing planet) in the two particular cases of Jupiter perturbed by Saturn (case 1) and Mars

[1]Mathematician, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts; regularly with Faculty of Sciences, Montpelier University, France.

perturbed by the earth (case 2). A comparison of my results with those of Le Verrier (1876) (case 1) and those of Clemence (1949) (case 2) will be carried out at a later time.

c. My results concerning the elimination of the short-periodic terms are developed in detail in Meffroy (1966a).

B.

a. The second step in the building of the theory is the elimination of the long-periodic terms, starting from T. In Meffroy (1966b), I indicated a direct solution of T. I pointed out an integral, ω = constant, of T. If $\omega \neq 0$, the linear variables are constants or linear functions of the time t; the angular variables are obtained through three quadratures. If $\omega = 0$, the linear variables are constants or linear functions of the expression $(t - t_0)^{2/3}$, t_0 being an arbitrary value of t; the angular variables are obtained through two quadratures, one of them elementary.

b. The determining function that eliminates the long-periodic terms is obtained by a linear partial differential equation of the first order whose coefficients are not constant, as in the case of the elimination of the short-periodic terms. This, then, is a very different analytical problem.

REFERENCES

LE VERRIER, U. J.

1876. Théorie du movement de Jupiter. Ann. Obs. Paris, Mém, vol. 11, pp. 105-272.

CLEMENCE, G. M.

1949. First order theory of Mars. Astron. Papers, Amer. Ephem. Naut. Almanac, vol. 11.

MEFFROY, J.

1966a. On Von Zeipel's method in general planetary theory. Smithsonian Astrophys. Obs. Spec. Rep. No. 229, 80 pp.

1966b. Sur un mode de résolution direct des équations canoniques résultant de l'élimination des termes à courte période d'une théorie planétaire générale du premier ordre. C. R. Acad. Sci. Paris, Ser. A, vol. 263, pp. 145-148.