NASA CR-167,839

NASA-CR-167839
19830014794

A Reproduced Copy

OF

NASA CR-167,839

Reproduced for NASA

by the

NASA Scientific and Technical Information Facility

NF01855

FFNo 672 Aug 65

BEST

AVAILABLE

COPY

(NASA-CR-167839) PROCEEDINGS OF THE NASA/MPRIA WORKSHOP: MATH/STAT Progress Report (Texas A&M Univ.) 183 p HC A09/MF A01 CSCL 09B G3/63

N83-23065 THRU N83-23072 Unclas 09827

DEPARTMENT OF MATHEMATICS

TEXAS A&M UNIVERSITY
COLLEGE STATION TEXAS

PROCEEDINGS OF THE

NASA/MPRIA WORKSHOP: MATH/STAT

Texas A&M University
College Station, Texas
January 27-28, 1983

Prepared for

Earth Resources Research Division
NASA/Johnson Space Center
Houston, Texas 77058

by

L. F. Guseman, Jr.
Principal Investigator
Department of Mathematics
Texas A&M University
College Station, Texas 77843

under

NASA Contract NAS 9-16664

"Mathematical Pattern Recognition
and Image Analysis Program"

TABLE OF CONTENTS

Introduction - L. F. Guseman, Jr. 1

Agenda 3

Participants and Other Attendees 5

Papers:

Estimating Proportions of Materials Using MIxture Models
Richard P. Heydorn and Rekha Basu 7

Nonparametric Probability Density Estimation for Data
Analysis in Several Dimensions - David W. Scott 37

Random Field Models For Use in Scene Segmentation
M. Naraghi 49

FUN.STAT and Statistical Image Representaions
Emanuel Parzen 65

Statistical Image Representations: Non-Gaussian
Classification - William B. Smith 77

Multivariate Time Series in Two Dimensions and the
Classificaion Problem - H. J. Newton 91

A Minimax Approach to Spatial Estimation Using Affinity
Matrices - Carl N. Morris 101

Localized Shrinkage Factors and Minimax Results
Hubert Kostal 109

Covariance Hypotheses for LANDSAT Data - Slide Presentation
Charles Peters 115

Covariance Hypotheses for LANDSAT Data
Henry P. Decell, Jr. and Charles Peters 125

A Hypothesis Test for the Rank of the Minimal Linear
Sufficient Statistic - Richard A. Redner and
William A. Coberly 135

An Adaptive Technique for Fitting LANDSAT Data
Larry L. Schumaker and L. F. Guseman, Jr. 151

Appendix 171

Fundamental Research Data Base 173

INTRODUCTION

by

L. F. Guseman, Jr.

The organizational meeting for the NASA Fundamental Research Program in "Mathematical Pattern Recognition and Image Analysis" (MPRIA) was held at the NASA/Johnson Space Center in August, 1982. At this meeting each of the fifteen principal investigators briefly outlined the goals of their particular proposed research efforts. Most of the efforts (those outside NASA) had just been funded (July 16, 1982), and investigations were just getting underway.

In order to gain a better understanding of and stimulate discussions between the individual research efforts, it was decided to conduct two technical workshops at Texas A&M University about six months into the program. The first workshop was held January 27-28, 1983 and consisted of investigators from the "Mathematics/Statistics" areas. The second workshop was held February 3-4, 1983 and consisted of investigators from the "Pattern Recognition" areas.

Each of the workshops was conducted in an informal manner. Most of the time was spent in lively technical discussions about each of the research efforts. Additional time was spent discussing the availability of data sets. Dr. R. P. Heydorn announced the availability of a data tape that has been compiled for use by the research teams. Details concerning the content and format of the tape are discussed in the document entitled "Fundamental Research Data Base" appearing in the Appendix of these proceedings.

Agendas and lists of participants for the workshops appear in their respective Proceedings.

This Page Intentionally Left Blank

NASA/MPRIA WORKSHOP: MATH/STAT

Texas A&M University
January 27-28, 1983
Room 510, Rudder Tower

Thursday, January 27:

8:00 - 8:30	Coffee and donuts
8:30 - 9:00	Overview: Fundamental Research Program R. P. Heydorn, NASA/JSC
9:00 - 10:15	Estimating Proportions of Materials Using Mixture Models R. P. Heydorn and R. Basu, NASA/JSC
10:15 - 10:30	Break
10:30 - 11:45	Some 3-D Density Estimates David Scott, Rice University
11:45 - 1:00	Lunch
1:00 - 2:15	Random Field Models for Use in Scene Segmentation Manouher Naraghi, JPL
2:15 - 2:30	Break
2:30 - 3:45	FUN.STAT and Statistical Image Representations Emanuel Parzen, W. B. Smith and H. J. Newton, TAMU
3:45 - 4:00	Break
4:00 - 5:15	A Minimax Approach to Spatial Estimation Using Affinity Matrices Carl Morris, UT Austin

Friday, January 28

8:00 - 8:30	Coffee and Donuts
8:30 - 9:45	Covariance Hypotheses for LANDSAT Data Charles Peters and H. P. Decell, Jr., University of Houston
9:45 - 10:15	Break

NASA/MPRIA WORKSHOP: MATH/STAT cont.

10:15 - 11:30	A Hypothesis Test for the Rank of a Linear Minimum Sufficient Statistic Richard Redner, University of Tulsa
11:30 - 1:00	Lunch
1:00 - 2:15	An Adaptive Technique for Fitting LANDSAT Data Larry Schumaker and L. F. Guseman, Jr., TAMU
2:15 - 4:00	Other Presentations, Discussion, and Symposium Planning

NASA/MPRIA WORKSHOP: MATH/STAT

January 27-28, 1983

Participants and Other Attendees:

Rekha Basu/SG3
NASA/Johnson Space Center
Houston, TX 77058

Raj Chhikara/C25
LEMSCO/Johnson Space Center
Houston, TX 77258

Henry P. Decell
Department of Mathematics
University of Houston
Houston, TX 77004

Alan Feiveson/SG3
NASA/Johnson Space Center
Houston, TX 77058

L. F. Guseman, Jr.
Department of Mathematics
Texas A&M University
College Station, TX 77843

Richard P. Heydorn/SG3
NASA/Johnson Space Center
Houston, TX 77058

J. R. Hill
Department of Mathematics
RLM 8-100
The University of Texas at Austin
Austin, TX 78712

Hubert Kostal
Department of Mathematics
RLM 8-100
The University of Texas at Austin
Austin, TX 78712

Wang Shu Lu
Department of Mathematics
RLM 8-100
The University of Texas at Austin
Austin, TX 78712

Carl Morris
Department of Mathematics
RLM 8-100
The University of Texas at Austin
Austin, TX 78712

Manouher Naraghi
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91103
Mail Stop 168-514

H. J. Newton
Institute of Statistics
Texas A&M University
College Station, TX 77843

Emanuel Parzen
Institute of Statistics
Texas A&M University
College Station, TX 77843

Charles Peters
Department of Mathematics
University of Houston
Houston, TX 77004

Richard Redner
Department of Mathematical Science
University of Tulsa
600 S. College Avenue
Tulsa, OK 74135

Larry L. Schumaker
Department of Mathematics and
Center for Approximation Theory
Texas A&M University
College Station, TX 77843

David Scott
Department of Mathematics Sciences
Rice University
P.O. Box 1892
Houston, TX 77251

N83
23066

UNCLAS

This Page Intentionally Left Blank

N83 23066

ESTIMATING PROPORTIONS OF MATERIALS USING MIXTURE MODELS

Richard P. Heydorn and Rekha Basu
NASA Johnson Space Center
Houston, Texas

1.0 INTRODUCTION

Let $F = \{f_\xi;\ \xi \epsilon R^N\}$ be a family of probability density functions and let G be a distribution function on R^N. For the given G we define a <u>mixture</u> density h as

$$h = \int f_\xi dG(\xi) \tag{1}$$

Since all the members of F are used in this definition, it makes sense to say that F defines a mapping, say $\mathcal{F}$, from the set of all G-distributions, say $\mathcal{G}$, to the set of all induced h-densities, say $\mathcal{H}$. If $\mathcal{F};\mathcal{G} \rightarrow \mathcal{H}$ is one-to-one and onto then we say that $\mathcal{H}$ is <u>identifiable</u>. This formulation is essentially due to Teicher (1). Thus, identifiability implies that, for a given mixture density h, a knowledge of the family F will allow us to uniquely determine G. This has practical implications for estimating the proportion of a material class on the ground using remotely sensed observations of that material. To illustrate the point, we offer the following example.

Suppose we are given spectral measurements, x, of points (pixels) on the ground which have been obtained from a satellite-multispectral scanner system. We imagine that these x's are observations on some

PRECEDING PAGE BLANK NOT FILMED

random variable X distributed according to density h. Suppose that through experimentation we have found that a given material class on the ground gives rise to measurements that are normally distributed as $N(\cdot;\mu,\sigma)$ but that the means μ, and variances change from region-to-region or from year-to-year. We know that in a given region there is a finite number of material classes, and that these classes can be described by members of a normal family; however, the means and variances are unknown. The mixture model that applies to this case is

$$h(x) = \sum_{j=1}^{M} \lambda_j \frac{1}{\sqrt{2\pi\sigma_j^2}} e_j^{-(1/2)(x-\mu_j)^2/(\sigma_j^2)} \quad (2)$$

where in this example G assigns a point probability λ_j to the points (μ_j, σ_j), $j=1,2,\ldots,M$. This is an example of a finite mixture model. Since the M material classes are associated with the parameters (μ_j, σ_j), $j=1,2,\ldots,M$, λ_j can be considered as the a-priori probability of observing the j-th class or λ_j is the proportion of the j-th class present in the given region. The primary aim is to determine the λ_j-values but to do that one has to estimate M, μ_j, σ_j, $j=1,2,\ldots,M$. Studies within the AgRISTARS program suggest that a multivariate version of the model given in equation (2) fits reasonably well to agricultural data, c.f. Lennington et al. (2) as well as to data from national vegetative classes. In those studies maximum likelihood estimation methods were used to estimate the λ_j's, the means, and the covariances. The number of classes, M, was determined by applying a heuristically derived algorithm.

There is yet another point to be made about the use of a mixture model for this application. Given a sequence of unlabeled observations $x_1,x_2,\ldots$, the proportions $\lambda_1,\lambda_2,\ldots,\lambda_M$ can presumably be determined. However, since the observations are unlabeled one cannot associate a class name to a given λ_j-value. Thus, there is a labeling problem associated with the use of the mixture model.

Two possible approaches to this problem are suggested. If one can obtain a random sample of labeled observations, then it would be possible to consider approaches aimed at testing the hypothesis.

"λ_j belongs to material class k" for combinations of k and j. Since ground enumeration methods are often required to obtain labeled observations, at least for foreign applications, this approach may be infeasible.

Another approach to the problem is to use observations on some auxiliary random variable to predict the mean behavior of the x-observations. Thus, one might attempt to use the auxiliary variable to predict the mean for the class of interest, and then from equation (2) associate the μ_j which is closest to that prediction. If y is the auxiliary random variable, then a better approach might be to consider the bivariate mixture model

$$h(x,y) = \sum_{j=1}^{M} \lambda_j f_{\xi_j}(x,y)$$

Knowing f_{ξ_j}, one can determine the regression function values

$$E_{\xi_j}(X|Y = y) = \int x f_{\xi_j}(x|y)dx, \quad j=1,2,\ldots,M$$

For the class of interest, it is often possible to establish the regression function, say $E_1(X|Y)$, having historical observations on X and Y from areas similar to the one being observed. Thus picking a λ_j that is associated with the class of interest is done by matching the regression function $E_1(X|Y)$ to one $E_{\xi_j}(X|Y)$, $j=1,2,\ldots,M$.

2.0 STUDY OBJECTIVES

We intend to pursue studies which are aimed at developing an approach to proportion estimation based on the notion of a mixture model. Specific objectives are to:

a) select appropriate parametric forms for a mixture model that appears to fit observed remotely sensed data,

b) develop methods for estimating the parameters in these models,

c) develop methods for labeling proportion determination from the mixture model,

and as a possible fourth objective

d) explore methods which use the mixture model estimates as auxiliary variable values in some proportion estimation scheme.

This latter objective admits the possibility that the λ_j-determinations may be only rough approximations to actual proportion, but are nevertheless useful as part of some other estimation scheme.

We have begun our studies by working on objective b) using the normal model form of equation (2). Our main purpose in mind is to develop methods for estimating M, since, in our opinion, least is known about estimating this parameter compared to the λ_j-values, the means, and the covariances. Interestingly, the approach we are pursuing also leads naturally to an estimate of the means.

3.0 Szegö's Solution to the Trigonometric Moment Problem — The Case of Equal and Known Variances

Consider a simple version of the mixture model in equation (2) in which the variances are all equal to σ and σ is known. Thus, we have

$$h(x) = \sum_{j=1}^{M} \lambda_j \frac{1}{\sqrt{2\pi\sigma^2}} e_j^{-(1/2)(x-\mu_j)^2/(\sigma^2)}$$

Taking the Fourier transform, we have

$$H(\omega) = \sum_{j=1}^{M} \lambda_j e^{i\omega\mu_j} e^{-(1/2)\omega^2\sigma^2}$$

Thus

$$D(\omega) = e^{\omega^2\sigma^2/2} H(\omega) = \sum_{j=1}^{M} \lambda_j e^{i\omega\mu_j}$$

Given the ω-values $\omega_k = \frac{k}{\omega_o}$ and letting $d_k = D(\omega_k)$ we have a representation of the complex numbers d_k as

$$d_k = \sum_{j=1}^{M} \lambda_j (e^{i\mu_j})^k. \qquad k=1,2,\ldots,n \tag{3}$$

Carathecdory proved that given the complex numbers $d_1, d_2, \ldots, d_n$ where $d_k \neq 0$ for some k, there exists an integer M, $1 < M < n$, and constants $\lambda_j e^{i\mu_j}$ such that $\lambda_j > 0$ and $\mu_j \neq \mu_\ell$, $\ell \neq j$, and the representation of equation (3) holds and is unique. The problem of determining M, λ_j, μ_j, $j=1,\ldots,M$ given the complex numberd d_k, $k=1,2,\ldots,n$ is called the trigonometric moment problem. Notice that the uniqueness of the representation is a consequence of the identifiability of normal mixtures.

The piont of interest for this study is the proof of the Caratheodory Theorem as given in Grenander and Szegö (3), since the proof gives a method for determining M and μ_j, j=1,2,...,M. To the best of our knowledge the proof is due to Szegö. We now sketch some of the ideas of the proof which will be of interest to us.

Given the complex numbers d_k we construct the Hermitian matrix

$$\mathbb{D} = \begin{pmatrix} 1 & d_1 & d_2 \cdots\cdots & d_n \\ d_{-1} & 1 & d_1 \cdots\cdots & d_{n-1} \\ \cdot & \cdot & \cdot & \cdot \\ d_{-n} & \cdot & \cdot\cdots\cdots & 1 \end{pmatrix}$$

where $d_{-k} = \bar{d}_k$; i.e., the complex conjugate of d_k.

For the representation in equation (3) $\mathbb{D}$ can be expressed as

$$\mathbb{D} = \sum_{j=1}^{M} \lambda_j \begin{pmatrix} 1 & e^{i\mu_j/\omega_o} & e^{2i\mu_j/\omega_o} \cdots & e^{ni\mu_j/\omega_o} \\ e^{-i\mu_j/\omega_o} & 1 & e^{i\mu_j/\omega_o} \cdots\cdots & e^{(n-1)i\mu_j/\omega_o} \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ e^{-ni\mu_j/\omega_o} & \cdots\cdots\cdots\cdots\cdots & & 1 \end{pmatrix}$$

which can be written as:

$$\mathbb{D} = \sum_{j=1}^{M} \lambda_j \begin{pmatrix} 1 \\ e^{-i\mu_j/\omega_o} \\ \cdot \\ \cdot \\ e^{-ni\mu_j/\omega_o} \end{pmatrix} \left(1, e^{i\mu_j/\omega_o}, \ldots, e^{ni\mu_j/\omega_o}\right) \tag{4}$$

Thus $\mathbb{D}$ is a linear combination of M ra·': 1 matrices and so its rank will not exceed M, and in fact is M. Certainly M<n.

Also from equation (4) we see that any Toeplitz form, $\upsilon' \mathbb{D}\upsilon$ is of the form

$$\upsilon' \mathbb{D}\upsilon = \sum_{j=1}^{M} \lambda_j \left| \sum_{\nu=o}^{M} \upsilon_\nu \, (e^{i\mu_j/\sigma_o})^\nu \right|^2$$

Hence, if υ are eigenvectors of $\mathbb{D}$ and $\upsilon' \mathbb{D}\upsilon$ is a (say the first) zero eigenvalue then

$$0 = \sum_{j=1}^{M} \lambda_j \left| \sum_{\nu=o}^{M} \upsilon_\nu \, (e^{i\mu_j/\sigma_o})^\nu \right|^2$$

Since $\mathbb{D}$ is a (n+1)x(n+1) matrix of rank less than or equal to n it must have at least one zero eigenvalue. Thus, it must be that the complex polynomial

$$P(Z) = \sum_{\nu=o}^{M} \upsilon_\nu Z^\nu$$

(assuming $\lambda_j > 0$, $j=1,2,\ldots,M$) has roots $Z_j = e^{i\mu_j/\sigma_o}$.

Summarizing the main points we have that:

a) The rank of $\mathbb{D}$ is the number of components in our normal mixture model.

b) The roots of the complex polynomial P(Z) lead to the means in our normal mixture model.

In passing we remark that in the study of time series Pisarenko (4) applied Szegö's approach. Refinements of the approach have also been proposed by Reddy et al. (5).

4.0 The Case of Unknown and Unequal Variances

The above approach makes use of the following fact about Fourier transforms. If $g(\cdot+y)$ is a translate of g and g has the Fourier transform $G(\omega)$ then the Fourier transform of its translate is $G(\omega)\cdot e^{-i\omega y}$. Hence the above approach attempts to find the number of translations and their amounts; but, to do that, h must be transformed so that only the translates are retained in the Fourier transform.

We pursue this general approach in considering the problem of unknown and unequal variances. We shall only sketch the main ideas of our approach.

For convenience, we assume that $\int_{-\infty}^{\infty} xh(x)dx = 0$ and define the truncation of h to be

$$h_b(x) = h(x)U_{(-b,b)}(x)$$

where

$$U_{(-b,b)}(x) = \begin{cases} 1, & x \in (-b,b) \\ 0, & x \notin (-b,b) \end{cases}$$

Since

$$\begin{aligned} U_{(-b,b)}(x) &= U_{(-b-\mu_j,b-\mu_j)}(x-\mu_j) \\ &= U_{(-b-\mu_j,b+\mu_j)}(x-\mu_j) - U_{(b-\mu_j,b+\mu_j)}(x-\mu_j) \\ &= U_{(-b-\mu_j,b+\mu_j)}(x-\mu_j) - U_{(-\mu_j,\mu_j)}(x-\mu_j-b) \end{aligned}$$

$$h_b(x) = \sum_{j=1}^{M} \lambda_j \frac{1}{\sqrt{2\pi\sigma^2_j}} e^{-(1/2)(x-\mu_j)^2/(\sigma^2_j)} U_{(-b-\mu_j,b+\mu_j)}(x-\mu_j)$$

$$- \sum_{j=1}^{M} \lambda_j \frac{1}{\sqrt{2\pi\sigma^2_j}} e^{-(1/2)(x-\mu_j)^2/(\sigma^2_j)} U_{(\mu_j,\mu_j)}(x-\mu_j-b)$$

or if we let

$$r_{bj}(x) = \frac{1}{\sqrt{2\pi\sigma^2_j}} e^{-(1/2)\frac{x^2}{\sigma^2_j}} U_{(-b-\mu_j,b+\mu_j)}(x)$$

$$s_{bj}(x) = \frac{1}{\sqrt{2\pi\sigma^2_j}} e^{-(1/2)\frac{x^2}{\sigma^2_j}} U_{(-\mu_j,\mu_j)}(x-b)$$

We see that h_b is a mixture of the translates of r_{bj} and s_{bj} which are truncated functions.

To retain only the "translation information" we will make use of the Shannon Sampling Theorem.

THEOREM (Shannon)

Let g be a function for which $g(x) = 0$ for $x \in (-\infty,-a)U[a,\infty)$. Then

$$G(\omega) = \sum_{r=-\infty}^{\infty} G(\frac{-k}{2a}) \frac{\sin(2a\omega - k)}{2a\omega - k}$$

We can rewrite this transform as

$$2i\omega G(\omega) = \left(\frac{1}{2a}\sum_{k=-\infty}^{\infty} C_k\, e^{-ik}\right) e^{i2a\omega}$$

$$- \left(\frac{1}{2a}\sum_{k=-\infty}^{\infty} C_k\, e^{ik}\right) e^{-i2a\omega}$$

$$+ \frac{i}{a}\sum_{k=-\infty}^{\infty} kC_k\, \frac{\sin(2a\omega - k)}{2a\,\omega - k}$$

where

$$C_k = G\left(\frac{-k}{2a}\right).$$

Lemma

Let $\ell(\omega) = \sum_{k=-\infty}^{\infty} kC_k \frac{\sin(2a\omega - k)}{2a\,\omega - k}$. If

$|\Sigma\, C_k| < \infty$, $|\Sigma\, kC_k| < \infty$ then $\lim_{\omega\to\infty} \ell(\omega) = 0.$

The Shannon Sampling Theorem plus the above lemma suggests that $2i\omega H_b(\omega)$ contains just the translation information for large values of ω. Indeed it can be shown that this is the case. The main result is finally stated in the following theorem.

Theorem

Let
$$V(\omega) = \sum_{j=1}^{M}\sum_{t=1}^{4} \alpha_{tj}\, e^{i\omega\delta_{tj}}$$

where $|\alpha_{tj}| > 0$ and δ_{tj} are real for $j=1,2,\ldots,M$, $t=1,2,3,4$. Let $\omega_\ell = \ell+2\pi N$, ℓ and N integers. Then for ω>M the matrix

$$B = \begin{pmatrix} V(\omega_0), & V(\omega_1), & \dots, & V(\omega_n) \\ V(\omega_{-1}), & V(\omega_0), & \dots, & V(\omega_{n-1}) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ V(\omega_{-n}), & \dots & \dots, & V(\omega_0) \end{pmatrix}$$

has rank 4M provided the constants δ_{tj} are distinct. Moreover, $B + (\bar{B})^{\circ}$ is Hermitian of rank 4M (where "-" denotes conjugate of and "o" denotes transpose of).

Thus our approach is to build a Hermitian matrix from $2iwH_b(\omega)$, as in the above theorem, let N get large, and apply Szegö method to compute M and the means.

5.0 Possible Extensions and Comments

In the proceeding approach h can be written as the convolution

$$h(x) = \sum_{j=1}^{M} \lambda_j \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\alpha_j^2}} e^{-(1/2)(x-y)^2/(\alpha^2{}_j)} \frac{1}{\sqrt{2\pi\beta^2}} e^{-(1/2)(y-\mu_j)^2/\beta^2} dy$$

where $\alpha_j^2 + \beta^2 = \sigma_j^2$. The first term in the integrand represents the "exponential decay" part of the mixture and the second the translation part. Our approach required that we essentially eliminate the contribution of the first term and preserve the contribution of the second. To do this, we made use of the exponential characteristics of the first term and the pure translation part of the second term. Thus, it would appear that representations that depend only on these two properties could also be approached by the above method. In particular one could consider a representation of the form

$$h(x) = \sum_{j=1}^{M} \lambda_j \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-(1/2)(x-y)^2/(\sigma^2{}_j)} g_j(y-\mu_j)dy$$

where $g_j(y-\mu_j) = \begin{cases} \gamma e^{-\gamma(y-\mu_j)}, & y>\mu_j \\ 0, & \text{otherwise} \end{cases}$

which would give skewed components in the mixture.

Finally, we recognize that the methods considered are most easily executed in one dimension. To handle the multidimensional problem, it may be possible to:

a) through transformations, develop vector valued random variables that have independent components. Then h could be considered as a product of marginal distributions.

b) consider conditional mixtures. That is consider e.g. h $(x|y)$ and solve the mixture problem for several fixed values of y.

c) treat only the marginal distributions and consider cases where this approach provides at least a good estimate of M. For such an approach one may be able to consider projections of the measurements which would attempt to bring out the true value of M.

REFERENCES

1. Teicher, H. "Identifiability of Finite Mixtures," Annuals of Mathematical Statistics, 34:1265-1269, 1963.

2. Lennington, R.K., C.T. Sorensen, R.P. Heydorn, "Can Crop Types be Resolved Using Mixture Distribution Components—Some Initial Results and Implications." Proceedings of 1982 Machine Processing of Remotely Sensed Data Symposium.

3. Grenander, U., G. Szegö, Toeplitz Forms and Their Applications, Berkeley, University of California Press. (1958).

4. Pisarenko, V.F. "The Retrieval of Harmonics from a Covariance Function," Geophys. J.R. Astr. Soc. (1973) 33:347-366.

5. Reddy, V.U., B. Egardt, T. Karilath, "Least Squares Type Algorithm for Adaptive Implementation of Pisarenko's Harmonic Retrieval Method," IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-30(3), June 1982.

MIXTURE MODELS

LET

$$\mathcal{F} = \{ f_\xi : \xi \in \mathbb{R}^N \}$$

FOR A DISTRIBUTION FUNCTION $G \in \mathcal{G}$
DEFINE A MIXTURE DENSITY h AS:

$$h = \int f_\xi \, dG(\xi) \qquad \text{.....} \quad (*)$$

LET $\mathcal{H}$ BE THE SET OF ALL INDUCED DENSITIES h, AND $\tilde{\mathcal{F}}$ THE MAPPING DEFINED IN (*):
THEN $\mathcal{H}$ IS IDENTIFIABLE IF

$$\tilde{\mathcal{F}} : \mathcal{G} \longrightarrow \mathcal{H}$$

IS 1-1 AND ONTO.

$\mathcal{H}$ IS A FINITE MIXTURE IF EACH G
ASSIGNS POSITIVE PROBABILITY MASS TO A
FINITE NUMBER OF ξ-VALUES.

IMPLICATIONS OF THE FINITE MIXTURE MODEL TO PROPORTION ESTIMATION

EXAMPLE

$x \sim$ SPECTRAL OBSERVATIONS

$$h(x) = \sum_{j=1}^{M} \lambda_j \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2}\frac{(x-\mu_j)^2}{\sigma_j^2}}$$

HERE G ASSIGNS PROBABILITY λ_j TO (μ_j, σ_j)

IF

- (μ_j, σ_j), $j = 1, 2, \ldots, M$ ARE UNIQUELY RELATED TO REAL MATERIAL CLASSES
- AND THE SPECTRAL DISTRIBUTIONS ARE NORM

THEN λ_j IS THE PROBABILITY OF FINDING THE j^{th} CLASS OR λ_j IS THE PROPORTION OF THE j^{th} CLASS.

KNOWING h & $\mathcal{F}$ $\Rightarrow$ KNOWLEDGE OF PROPORTION

i.e., $\mathcal{F}^{-1}: \mathcal{H} \rightarrow \mathcal{G}$

Segment Number	Ground Truth Proportion (%)	Direct Proportion Estimate (%)
1544	26.81	26.40
1394	41.48	39.57
1650	13.73	10.70
1920	15.99	13.88
1636	50.16	50.42
1663	53.98	53.42
1676	7.06	0.0
1566	37.32	28.32
1899	67.51	59.03
1825	34.40	29.43

Avg. G.T. Prop. = 34.84

Bias = -3.75
Variance = 3.26

Relative Bias = -0.11
Coefficient of Variation = .0.09

Table 3. Proportion Estimates of Small Grains Obtained from the Mixture Model

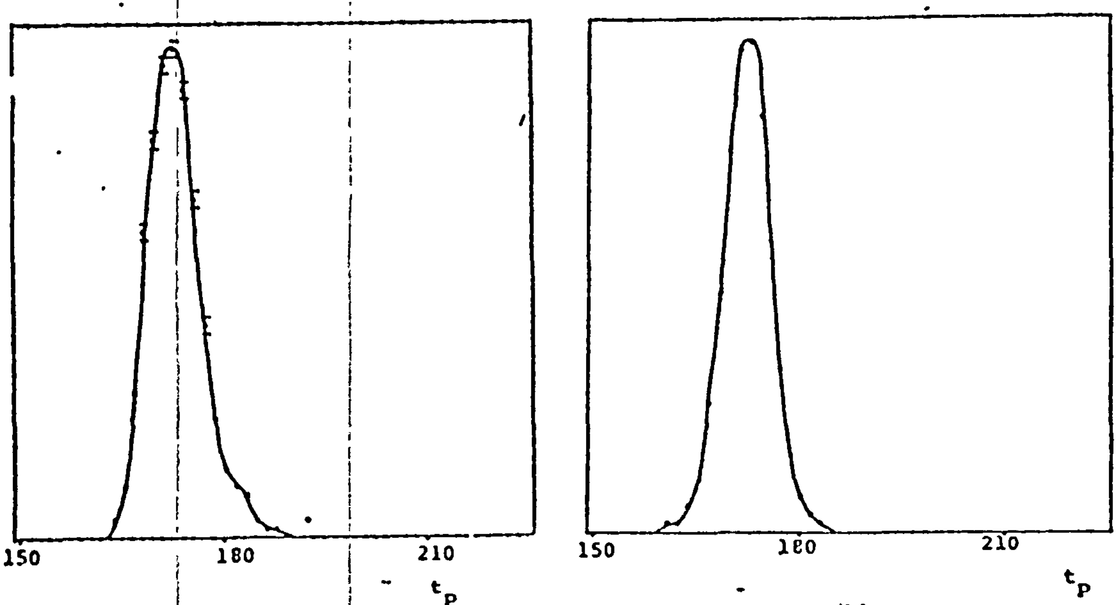


Figure 2. (a) Ground truth distribution for pure small grains pixels from segment 1899.
(b) CLASSY estimated distribution for small grains using all pure pixels from segment 1899.

THE LABELING PROBLEM

GIVEN OBSERVATIONS $x_1, x_2 \ldots$ WE CAN COMPUTE λ_j-VALUES. WE NEED TO ASSIGN MATERIAL CLASS NAMES TO THESE VALUES:

SPECTRAL CLASSES → MATERIAL CLASSES

HYPOTHESIS TESTING APPROACH

EXAMPLE

GIVEN $X_1, X_2, \ldots X_m$ FROM CLASS "k"

$$\frac{P_r(\text{SPECTRAL CLASS "j"} \mid X_1, X_2, \ldots, X_m)}{P_r(\text{NOT SPECTRAL CLASS "j"} \mid X_1, X_2, \ldots, X_m)} \longrightarrow C, \; m \to \infty$$

$C = \infty \Rightarrow$ j IS k

$C = 0 \Rightarrow$ j IS NOT k

"Prediction Model" Approach

X ~ (Transformed) Spectral Variable

Y ~ Auxiliary Variable

Prediction Models

$$E_k(X|Y) \quad , k = 1, 2, \ldots, K$$

Mixture Model

$$h(x,y) = \sum_{j=1}^{m} \lambda_j \, f_j(x,y)$$

Gives

$$E_j(X|Y=y) = \int x \, f_j(x|y) \, dx$$

Match Regression Functions

STUDY OBJECTIVES

FIND METHODS FOR:

a) PICKING f FROM REAL DATA

⇒ b) ESTIMATING MIXTURE MODEL PARAMETERS

c) LABELING PROPORTIONS

AND POSSIBLY

d) EXPLORE METHODS THAT USE MIXTURE MODEL ESTIMATES AS AUXILIARY VARIABLES IN SOME PROPORTION ESTIMATION SCHEME.

Szegö's Solution to the Trigonometric Moment Problem — The Case of Equal and Known Variances

MODEL

$$h(x) = \sum_{j=1}^{M} \lambda_j \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu_j)^2}{\sigma^2}}$$

FOURIER TRANSFORM (CHARACTERISTIC FUNCTION)

$$H(\omega) = \sum_{j=1}^{M} \lambda_j e^{i\omega\mu_j} e^{-\frac{1}{2}\omega^2\sigma^2}$$

LET

$$D(\omega) = e^{\omega^2\sigma^2/2} H(\omega) = \sum_{j=1}^{M} \lambda_j e^{i\omega\mu_j}$$

& FOR $\omega_k = \frac{k}{\omega_0}$ LET $d_k = D(\omega_k)$

$$d_k = \sum_{j=1}^{M} \lambda_j \left(e^{i\frac{\mu_j}{\omega_0}}\right)^k \quad \text{- - - —} \quad (*)$$

GIVEN $d_1, d_2, \ldots, d_m$ CARATHÉODORY PROVED the REPRESENTATION (*) is <u>UNIQUE</u>

A PROOF FOLLOWS FROM TEICHER'S THEOREM ON THE IDENTIFIABILITY OF NORMAL MIXTURES.

Szegö's Approach To The Proof of The Caratheodory Theorem.

Form The Hermitian Matrix

$$\mathbb{D} = \begin{pmatrix} 1 & d_1 & d_2 & \cdots & d_m \\ d_{-1} & 1 & d_1 & \cdots & d_{m-1} \\ \vdots & & & & \\ d_{-m} & \cdots & & \cdots & 1 \end{pmatrix}$$

Thus

$$\mathbb{D} = \sum_{j=1}^{M} \lambda_j \begin{pmatrix} 1 & e^{i\frac{\mu_j}{\omega_0}} & e^{2i\frac{\mu_j}{\omega_0}} & \cdots & e^{mi\frac{\mu_j}{\omega_0}} \\ e^{-i\frac{\mu_j}{\omega_0}} & 1 & e^{i\frac{\mu_j}{\omega_0}} & \cdots & e^{(m-1)i\frac{\mu_j}{\omega_0}} \\ \vdots & & & & \\ e^{-mi\frac{\mu_j}{\omega_0}} & \cdots & & \cdots & 1 \end{pmatrix}$$

OR

$$\mathbb{D} = \sum_{j=1}^{M} \lambda_j \begin{pmatrix} 1 \\ e^{-i\frac{\mu_j}{\omega_0}} \\ \vdots \\ e^{-mi\frac{\mu_j}{\omega_0}} \end{pmatrix} \begin{pmatrix} 1 & e^{i\frac{\mu_j}{\omega_0}} & \cdots & e^{mi\frac{\mu_j}{\omega_0}} \end{pmatrix}$$

$\Rightarrow$ Rank $\mathbb{D} \le M$

TOEPLITZ FORM $v'Dv = 0$ IF v IS THE EIGENVECTOR ASSOCIATED WITH (FIRST) ZERO EIGENVALUE. THUS

$$0 = \sum_{j=1}^{M} \lambda_j \left| \sum_{\nu=0}^{M} v_\nu (e^{i\mu_j})^\nu \right|^2$$

OR

$$P(z) = \sum_{\nu=0}^{M} v_\nu z^\nu$$

HAS ROOTS $z_j = e^{i\mu_j}$

THE CASE OF UNEQUAL AND UNKNOWN VARIANCES.

GIVEN, $h(x)$ & THE REPRESENTATION

$$h(x) = \sum_{j=1}^{M} \lambda_j \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2}\frac{(x-\mu_j)^2}{\sigma_j^2}}$$

FIND M, λ_j, μ_j, σ_j, $j=1,2,\ldots,M$

NOTE:

FOR THE CASE OF EQUAL VARIANCES WE MADE USE OF THE FOLLOWING PROPERTY OF FOURIER TRANSFORMS

$g_y = g(\cdot + y)$ A TRANSLATE OF g $\Rightarrow$

$$G_y(\omega) = G(\omega)\underbrace{e^{i\omega y}}_{\text{PURE TRANSLATE}}$$

WE CONSIDER AN APPROACH FOR FINDING M, μ_j, $j=1,2,\ldots,M$ THAT USES SZEGÖ'S METHOD.

MAIN IDEAS

ASSUME $\int x h(x) dx = 0$, AND LET

$$h_b(x) = h(x) U_{(-b,b)}(x)$$

$$U_{(-b,b)}(x) = \begin{cases} 1, & x \in (-b,b) \\ 0, & x \notin (-b,b) \end{cases}$$

NOW

$$U_{(-b,b)}(x) = U_{(-b-\mu_j,\, b-\mu_j)}(x-\mu_j) = U_{(-b-\mu_j,\, b+\mu_j)}(x-\mu_j) - U_{(b-\mu_j,\, b+\mu_j)}(x-\mu_j)$$

$$= U_{(-b-\mu_j,\, b+\mu_j)}(x-\mu_j) - U_{(-\mu_j,\, \mu_j)}(x-\mu_j-b)$$

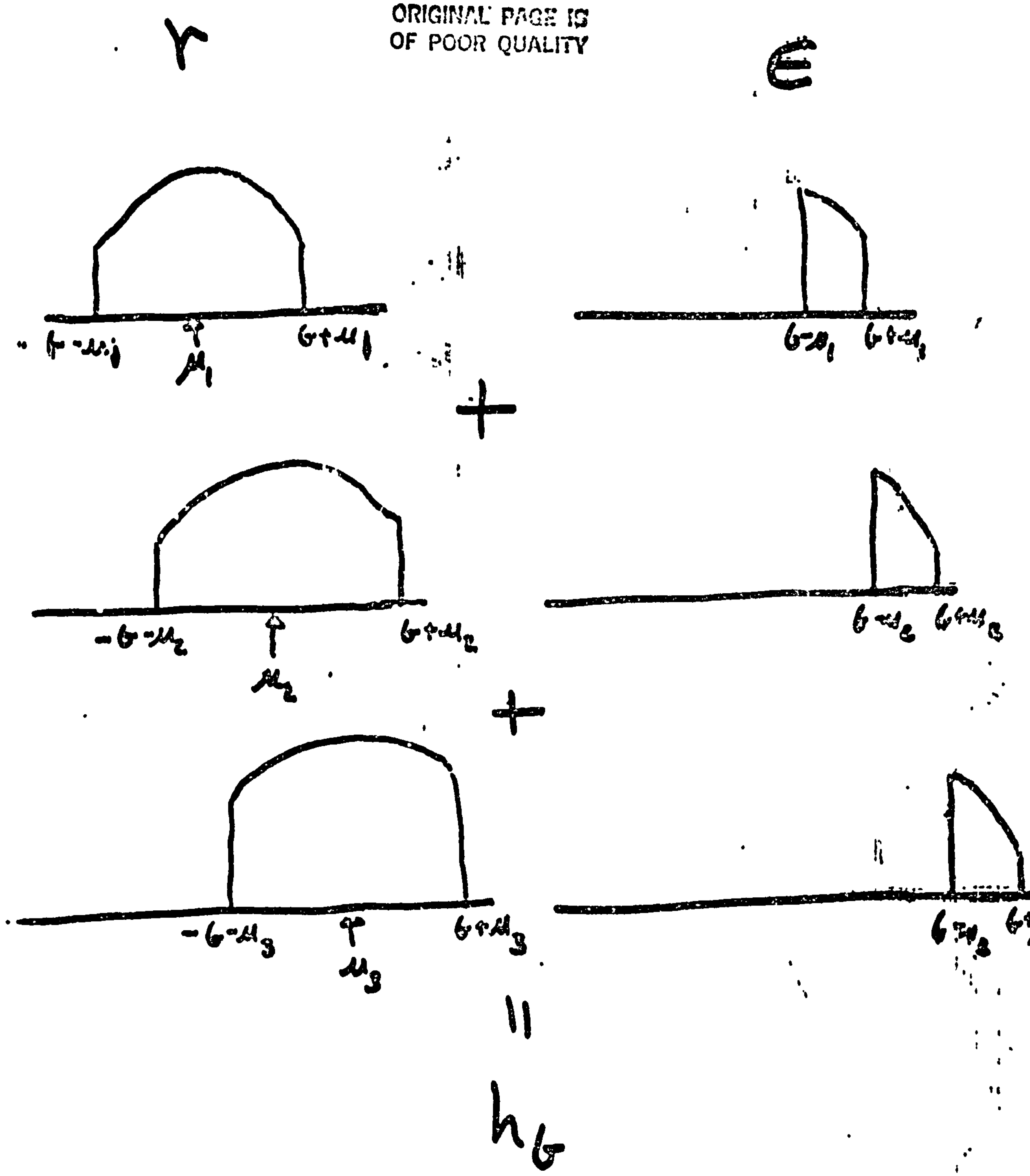
LET

$$r_{b_j}(x) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2}\frac{x^2}{\sigma_j^2}} U_{(-b-\mu_j,\, b+\mu_j)}(x)$$

$$\epsilon_{b_j}(x) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2}\frac{x^2}{\sigma_j^2}} U_{(-\mu_j,\, \mu_j)}(x-b)$$

THEN

$$h_b(x) = \sum^{M} \lambda_j r_{b_j}(x-\mu_j) - \sum^{M} \lambda_j \epsilon_{b_j}(x-\mu_j)$$



TO RETAIN ONLY THE "TRANSLATION INFORMATION" IN THE F, G FUNCTIONS WE MAKE USE OF

THEOREM (SHANNON)

LET g BE A FUNCTION FOR WHICH $g(x) = 0$ FOR $x \in (-\infty, -a] \cup [a, \infty)$. THEN

$$G(\omega) = \sum_{k=-\infty}^{\infty} G\left(\frac{-k}{2a}\right) \frac{\sin(2a\omega - k)}{2a\omega - k}$$

OR REWRITING G

$$2i\omega G(\omega) = \left(\frac{1}{2a} \sum_{k=-\infty}^{\infty} c_k e^{-ik}\right) e^{i2a\omega}$$

$$- \left(\frac{1}{2a} \sum_{k=-\infty}^{\infty} c_k e^{ik}\right) e^{-i2a\omega}$$

$$+ \frac{1}{a} \sum_{k=-\infty}^{\infty} k c_k \frac{\sin(2a\omega - k)}{2a\omega - k}$$

$$c_k = G\left(\frac{-k}{2a}\right).$$

LEMMA

LET $l(\omega) = \sum_{k=-\infty}^{\infty} k C_k \frac{\sin(2a\omega - k)}{2a\omega - k}$

IF $|\sum C_k| < \infty$, $|\sum k C_k| < \infty$ THEN

$\lim_{\omega \to \infty} l(\omega) = 0$.

MAIN RESULT:

THEOREM

LET $V(\omega) = \sum_{j=1}^{M} \sum_{\ell=0}^{4} d_{\ell j} e^{-i\omega \delta_{\ell j}}$

$|d_{\ell j}| > 0$, $\delta_{\ell j}$ REAL. LET $\omega_\ell = \ell + 2\pi N$, THEN FOR $m \geq M$

$$B = \begin{pmatrix} V(\omega_0), V(\omega_1), \dots, V(\omega_m) \\ V(\omega_{-1}) \; V(\omega_0), \dots, V(\omega_{m-1}) \\ \vdots \\ V(\omega_{-m}) \; \dots \dots \; V(\omega_0) \end{pmatrix}$$

HAS RANK $4M$ PROVIDED $\delta_{\ell j}$ ARE DISTINCT. ALSO $B + (\bar{B})'$ IS HERMITIAN OF RANK $4M$.

POSSIBLE EXTENSIONS

NOTICE THAT

$$h(x) = \sum_{j=1}^{M} \lambda_j \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi \alpha_j^2}} e^{-\frac{1}{2}\frac{(x-y)^2}{\alpha_j^2}}}_{\text{EXPONENTIAL DECAY}} \underbrace{\frac{1}{\sqrt{2\pi\beta^2}} e^{-\frac{1}{2}\frac{(y-\mu_j)^2}{\beta^2}}}_{\text{TRANSLATION}}$$

$$\alpha_j^2 + \beta^2 = \sigma_j^2$$

POSSIBLE ALTERNATIVE TO ACCOUNT FOR <u>SKEWED</u> DISTRIBUTIONS

$$h(x) = \sum_{j=1}^{M} \lambda_j \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2}\frac{(x-y)^2}{\sigma_j^2}} g_j(y-\mu_j)\, dy$$

$$g_j(y-\mu_j) = \begin{cases} \gamma e^{-\gamma(y-\mu_j)}, & y \geq \mu_j \\ 0, & \text{OTHERWISE} \end{cases}$$

Problems

Would Be Difficult To Extend To The Multidimensional Case.

Possibilities:

a) Develop Transformations of The Satellite observations That Lead To Independent Variables. Then h would Be A product of Marginals

b) Consider A Conditional Mixture, h(s|y)

c) Use The Marginals To Get A "Good" Estimate of M. Consider Projections of The Satellite observations.

N83
23067

UNCLAS

This Page Intentionally Left Blank

PRECEDING PAGE BLANK NOT FILMED

N83 23067

37

Nonparametric Probability Density Estimation for Data Analysis in Several Dimensions[1,2]

ORIGINAL PAGE IS OF POOR QUALITY

David W. Scott

Rice University
Houston, TX 77251

1. Introduction

Our purpose in this paper is to illustrate how nonparametric probability density estimates, in particular the corresponding contour curves, are a useful adjunct to scatter diagrams when performing a preliminary examination of a set of random data in several dimensions. For a preliminary approach we generally want to perform fairly simple tasks with free-form techniques to uncover structures and features of interest in the data. Such procedures are often graphical and unlike summary statistics seldom lead to much compression of the data. Tukey (1977) presents a wealth of such procedures. One which well illustrates the power and flexibility of these preliminary procedures is the running median smoothing algorithm for time series data (with resmoothing of the rough and the like). Other graphical techniques for multivariate data are presented in Tukey and Tukey (1981).

For preliminary viewing of one-dimensional data, both scatter diagrams and frequency curves such as histograms are widely and successfully employed to examine clustering, tail behavior, and skewness of

[1]This research was supported in part by the Army Research Office under DAAG-29-82-K-0014 and by NASA/Lockheed under PO-0200100079.

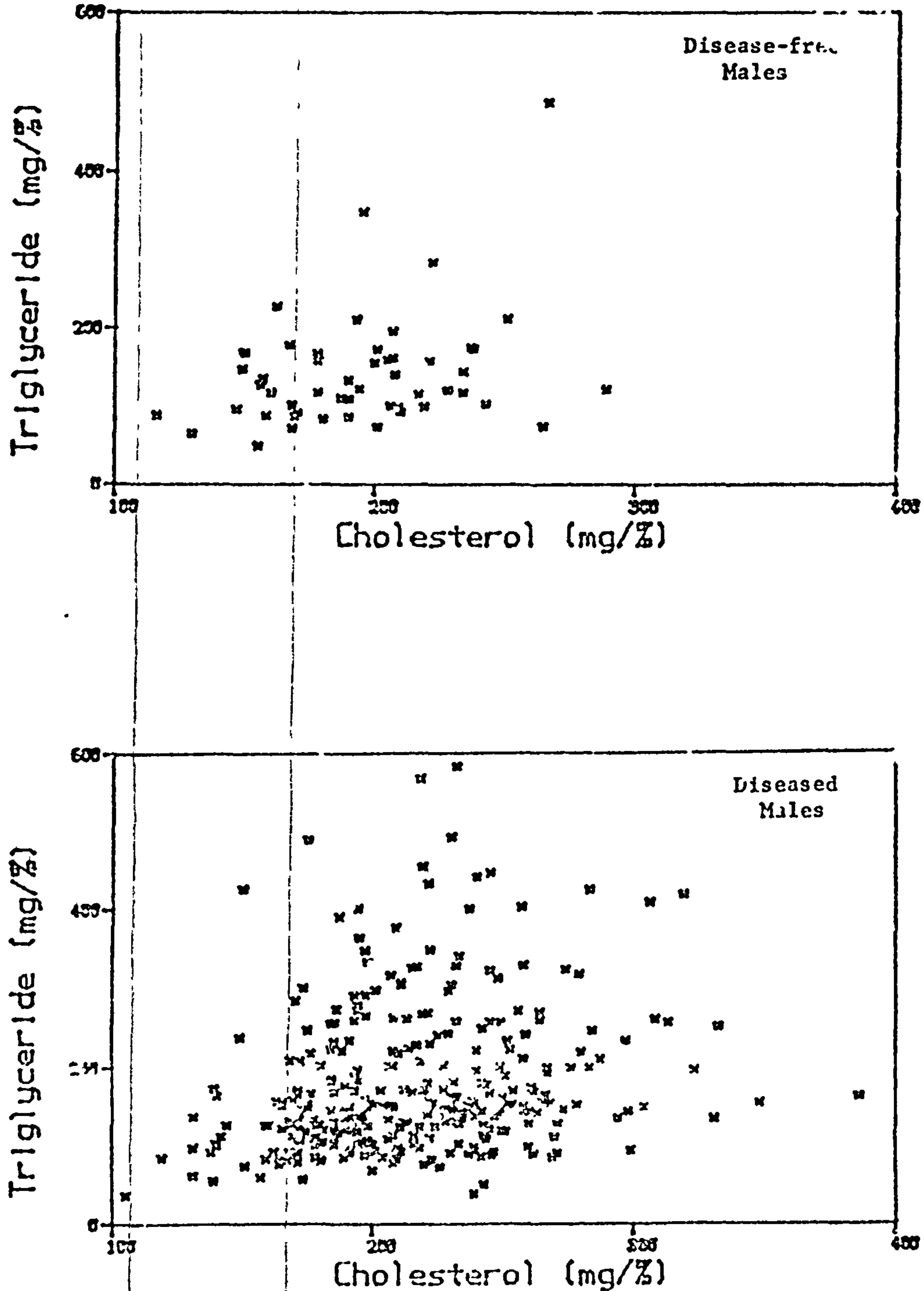
[2]To appear in the ... Design of Experiments Conference.

data. For bivariate data, scatter diagrams are in practice widely preferred to bivariate frequency curves. Scatter diagrams of three dimensional data may be realized by viewing a projection of the data on a rotating plane represented by the screen on a computer graphics terminal. For higher dimensions carefully selected projections may also be viewed, and sophisticated techniques have been developed, and are evolving, for choosing good projections (Friedman and Tukey, 1974). Apparently the success of frequency curves in one dimension has not readily extended to higher dimensions. It is an open question as to the number of dimensions that may be successfully visualized with a nonparametric density estimator under various conditions (sample size, for example). It is our purpose to illustrate the power of preliminary frequency curves as an adjunct to viewing scatter diagrams.

2. Bivariate Data

We shall examine a data set which contains information on the status of the coronary arteries of 371 men suspected of having heart disease, having experienced episodes of severe chest pain. These data have been more fully described and analyzed; see Gotto, *et al.* (1977) and Scott, *et al.* (1978). After visual examination of the coronary arteries by angiography, 51 men were determined to be free of significant coronary artery disease. It was of interest to compare the levels of blood fats, plasma cholesterol and plasma triglyceride concentrations, between the group of 51 disease-free males and the group of 320 diseased males. The scatter diagrams of these two data sets are displayed in Figure 1. Patients with elevated levels of cholesterol and

Figure 1. Scatter Diagrams



triglyceride are evident among the diseased males. This observation is difficult to evaluate in light of the large difference in sample sizes. However, it is unlikely that a larger sample of 320 disease-free males would result in a scatter diagram similar to that of the 320 diseased males.

To obtain a nonparametric density contour plot we computed a bivariate product kernel estimate (Epanechnikov, 1969) given by

$$f(x,y) = \frac{1}{nh_x h_y} \sum_{i=1}^{n} K\left(\frac{x_i - x}{h_x}\right) K\left(\frac{y_i - y}{h_y}\right) \quad (1)$$

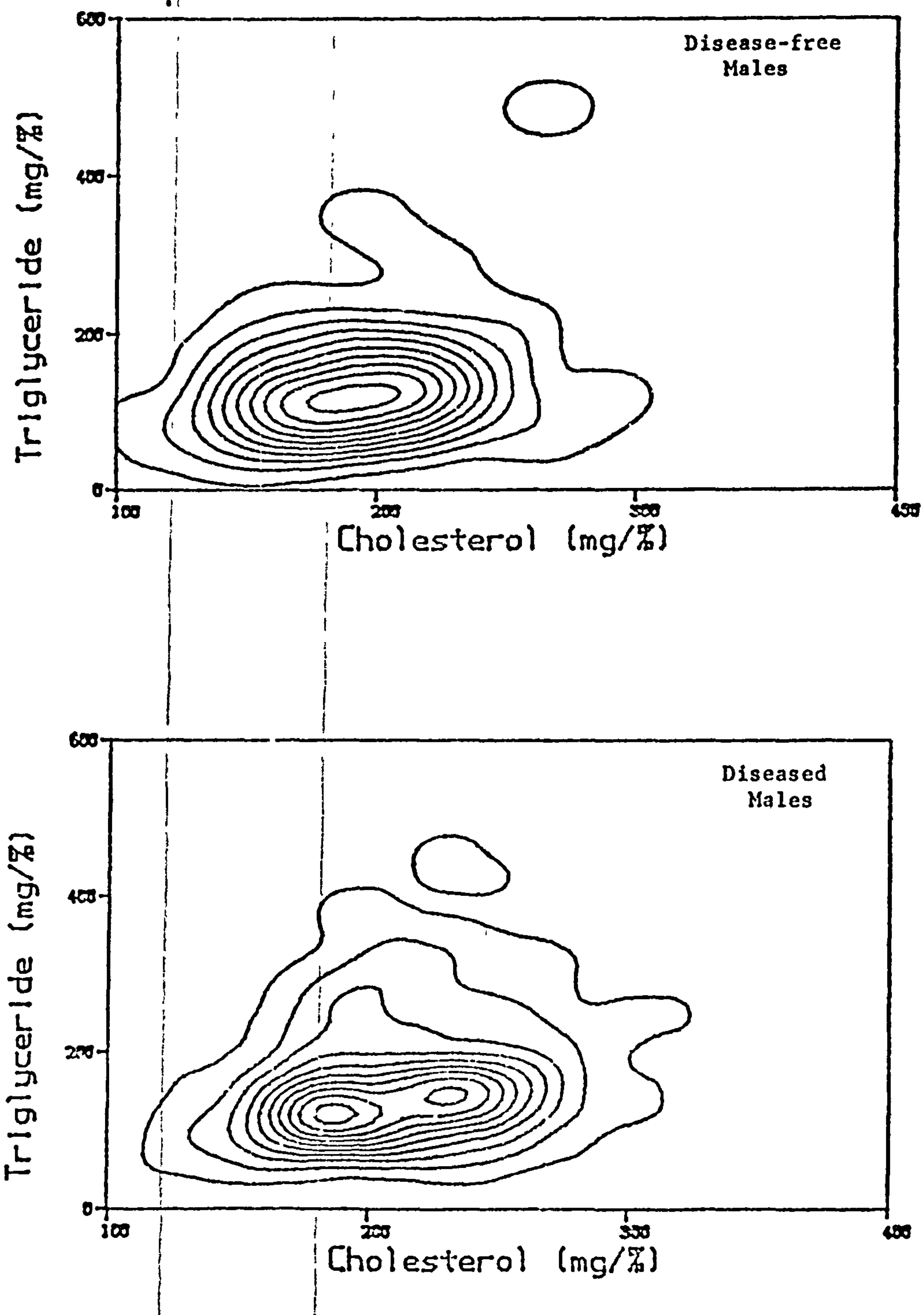
using a quartic (biweight) kernel

$$K(z) = \frac{15}{16}(1-z^2)^2 \, I_{[-1,1]}(z) \quad (2)$$

and preliminary values of the smoothing parameters given by $h_x = 2 s_x n^{-1/6}$ where s_x represents a trimmed and pooled estimate of the standard deviation for the two groups with a similar expression for h_y. Density values were computed over a grid of 150 by 90 points. When applied to the data for the diseased males, the contour plot reveals a striking bimodal feature, as shown in Figure 2. The contours of equal probability are at the ten levels 0.05 to 0.95 in increments of 0.10 as a fraction of the respective maximal modal levels. The density function of the disease-free males could be well approximated by a bivariate Normal form. Its mode coincides with the left of the two modes in the density function of the diseased males.

The contour plots have helped emphasize a feature in the scatter diagram that might have gone unnoticed. The contour plots also aid in compensating for the difference in sample sizes. The discovery of the bimodal feature led to formulation of a complex cholesterol-triglyceride

Figure 2. Bivariate Density Contours



interaction in the model for estimating the risk of coronary artery disease. Clinically, the difference of 50 mg/% between the two modes in Figure 2 for the diseased males is greater than the reduction in cholesterol by dietary intervention (which usually achieves proportional reductions in the range of 10 to 15 percent).

3. Trivariate Data

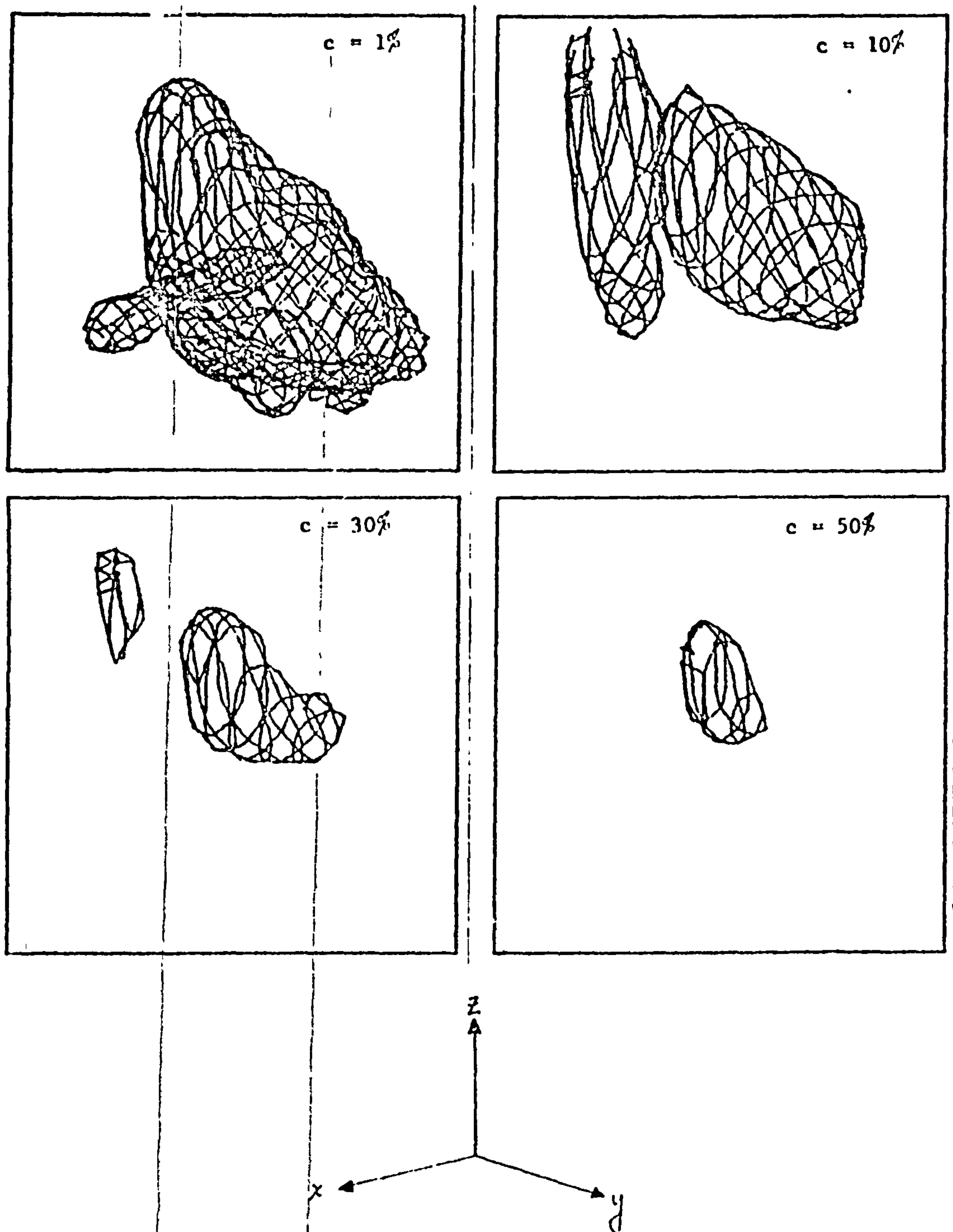
The data presented in this section were obtained by processing four-channel Landsat data measured over North Dakota during the summer growing season of 1977 and were furnished by Dick Heydorn of NASA/Houston and Chuck Sorensen of Lockheed/Houston. The sample contains approximately 21,000 points, each representing a 1.1 acre pixel, covering a 5 by 6 nautical mile section. On each pass over an individual pixel by the Landsat satellite, the four channel readings were combined into a single value that measures the "greenness" of the pixel at that time. The greenness of a pixel was plotted as a function of time from the five passes during the growing season. Finally, Badhwar's (1982) growth model was fitted to this curve. This model has three parameters which are contained in each trivariate data point. The first variable (x) gives the time the "crop" (if any) ripened. The second variable (y) measures the approximate time to ripen. And the third variable (z) measures the level of "greenness" at the time of ripening. Although it is natural to group these data by actual type of ground cover for classification procedures, we have not done so here.

It is not possible to present a satisfactory picture of a three-dimensional scatter diagram of these data for this article. However, on an AED512 terminal with 512 by 512 resolution, a projection of these data onto the screen typically displayed only 4000 points, the rest being "hidden" behind displayed points. Viewed from several different angles, various shapes and features in the data were easily perceived. Color was used to indicate the level of the variable perpendicular to the screen.

We can present density contours of an estimate f(x,y,z). Consider an equiprobable contour at level c; that is, consider those points (x,y,z) satisfying the equation f(x,y,z) = c. The solution of this equation for a smooth density estimate f is a smooth surface (or surfaces) in R^3. This surface may be displayed by intersecting it with a series of planes displaced equal distances along the co-ordinate axes, in the following, along only the x and y axes. In Figure 3, we display the surface for c = 1% of the maximal mode value. Comparing Figure 3 to the corresponding scatter diagram on the same projection plane reveals how surprisingly little of the data space is enclosed in this contour. In the scatter diagram our eyes focused on rays of points that seemed interesting but represented only a small fraction of the data. Also notable in Figure 3 is a cylindrical shape disjoint and behind the larger surface. This feature was also clearly visible in the scatter diagram and represents acres in which sugar beets were grown. Apparently the method by which sugar beets are harvested leads to a singularity in the estimation of the growth model parameters with $y \approx 0$.

Expanding the scale by a factor of 2 while retaining the same center as in the c = 1% picture, we show the contour shapes at levels c = 10%, 30%, and 50% of modal height. Notice how each contour shape

Figure 3. Trivariate Density Contours



"fits" inside the preceding one. Also observe how multimodal features appear in this space. Three modes are shown in this sequence. On a color graphics terminal, we may simultaneously view these and other contours by using different colors to draw each contour.

Again, the density plots have complemented and added to our understanding of these data. It is easier to see inside the data cloud with this representation and also makes rotation of the data cloud less important.

4. Computational Considerations

A new algorithm and density estimator were developed to display the trivariate contour plots and we hope to report on it in another paper (Scott, 1983b). Speed is an important factor in an interactive environment. The kernel method used in the bivariate case becomes excruciatingly slow when presented with 21,000 points in three dimensions. In real time, a few minutes were required on a Vax 11/780 to compute the bivariate kernel contours for 320 points on a 150 by 90 mesh. To generate the pictures in Figure 3, we evaluated the density on a 30 by 30 by 30 mesh for 21,000 points. A straightforward kernel estimator would have required several hours to compute!

The histogram estimator is extremely efficient computationally, but very inefficient statistically -- and relatively more inefficient in higher dimensions than kernel methods. One recent discovery indicates that the frequency polygon may be a good choice of a nonparametric density estimator since it is computationally equivalent to a histogram but statistically similar to a kernel estimate (Scott, 1983a). However, the frequency polygon in several dimensions suffers from sensitivity to choice of cell boundaries. The new algorithm addresses this problem and is asymptotically equivalent to a certain kernel estimate. Other fast preliminary estimates in one and two dimensions may be obtained by numerical approximation of kernel estimates in place of statistical approximation, which we prefer.

5. Where Do We Go?

We do not really know for how many dimensions nonparametric density estimates will be useful and feasible. Scatter diagrams have been used in a highly interactive environment to visualize nine-dimensional data (Tukey, Friedman, and Fisherkeller, 1976). Many possible strategies may be envisioned for using color and motion to examine data in more than three dimensions. We expect much progress in this area. But for larger and larger data sets requiring sophisticated analysis, we believe that density-based methods will be both efficient and effective.

ORIGINAL PAGE IS
OF POOR QUALITY

REFERENCES

Badhwar, G.D., J.G. Carnes, and W.W. Austin (1982), "Use of Landsat-Derived Temporal Profiles for Corn-Soybean Feature Extraction and Classification," Remote Sensing of Environment, 12, 57-79.

Epanechnikov, V.A. (1969), "Nonparametric Estimates of a Multivariate Probability Density," Th. Prob. and Appl., 14, 153-158.

Friedman, J.H. and J.W. Tukey (1974), "A Projection Pursuit Algorithm for Exploratory Data Analysis," IEEE Trans. Comp. C-23, 881-890.

Gotto, A.M., G.A. Gorry, J.R. Thompson, J.S. Cole, R. Trost, D. Yeshurun, M.E. DeBakey (1977), "Relationship Between Plasma Lipid Concentration and Coronary Artery Disease in 496 Patients," Circulation, 56:5, 875-883.

Scott, D.W. (1983a), "Optimal Frequency Polygons: Theory and Application," submitted.

Scott, D.W. (1983b), "Average Shifted Histograms," working paper.

Scott, D.W., A.M. Gotto, J.S. Cole, and G.A. Gorry (1978), "Plasma Lipids as Collateral Risk Factors in Coronary Artery Disease - A Study of 371 Males with Chest Pain," J. Chronic Diseases, 31, 337-345.

Tukey, J.W. (1977), Exploratory Data Analysis, Addison-Wesley, Reading, MA.

Tukey, J.W., J.H. Friedman, and M.A. Fisherkeller (1976), "PRIM-9, an Interactive Multidimensional Data Display and Analysis System," Proc. 4th Inter. Congress for Stereology, Sept. 4-9, 1975, Gaithersburg, Maryland.

Tukey, P.A. and J.W. Tukey (1981), "Graphical Display of Data Sets in 3 or More Dimensions," in Interpreting Multivariate Data, V. Barnett, ed., John Wiley & Sons, New York.

This Page Intentionally Left Blank

OMIT TO P.65

ORIGINAL PAGE IS OF POOR QUALITY

Random Field Models For Use in Scene Segmentation

M. Naraghi

Jet Propulsion Laboratory
California Institute of Technology

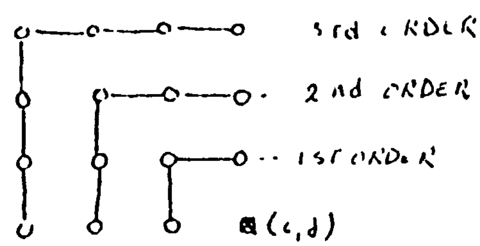
1 – PRELIMINARIES & NOTATION:

For a single band image let there be M classes $\omega_1, \dots, \omega_M$ where the pixels in each class are members of a 2-D Gaussian and stationary random process with known apriori means $\mu_1, \dots, \mu_M$ and autocorrelations $R_1(k,l), \dots, R_M(k,l)$ so for k^{th} class

(1)
$$\mu_k = E\, I^k(i,j)$$
$$R_k(|m-i|,|n-j|) = E\left[I^k(m,n)-\mu_k\right]\left[I^k(i,j)-\mu_k\right]$$

where $I^k(i,j)$ denotes the intensity at pixel (i,j) in the k^{th} class and E is the expectation operator.

On a 2-D grid let the autoregressive order associated with a location (i,j) be defined as shown



2- Autoregressive Models:

$$x(i,j) = \sum_{\substack{k=0 \\ k \neq l \neq 0}}^{P} \sum_{l=0}^{P} \alpha_{kl}\, x(i-k, j-l) + \sigma\, u(i,j)$$

$$E\, u(i,j) = 0$$

$$E\, u(i,j)\, u(k,l) = \begin{cases} 1 & \text{if } k=i \text{ \& } j=l \\ 0 & \text{otherwise} \end{cases}$$

α_{kl} are constants if $x(m,n)$ is stationary.

otherwise α_{kl} are a function of (i,j).

Example of such a model is

$$x(i,j) = \alpha_1 x(i, j-1) + \alpha_2 x(i-1, j) + \alpha_3 x(i-1, j-1) + \sigma\, u(i,j)$$

which corresponds to the correlation

$$R(\tau_1, \tau_2) = \sigma_s^2\, e^{-\delta_1|\tau_1| - \delta_2|\tau_2|}$$

$$\alpha_1 = e^{-\delta_2}$$

$$\alpha_2 = e^{-\delta_1}$$

$$\alpha_3 = -e^{-\delta_1 - \delta_2} = -\alpha_1\alpha_2$$

$$\sigma^2 = \sigma_s^2 (1 - \alpha_1^2 - \alpha_2^2 + \alpha_1^2\alpha_2^2)$$

Modelling Procedure :

For a given $R(m,n)$ and an autoregressive order P, the model will be

$$x(i,j) = \sum_{\substack{k=0 \\ k+l\neq 0}}^{P} \sum_{l=0}^{P} \alpha_{kl}\, x(i-k, j-l) + \sigma\, u(i,j)$$

α_{kl} and σ are found such that

$$E\left[\sigma\, u(i,j)\right]^2 = E\left[x(i,j) - \sum_{\substack{k=0 \\ k+l\neq 0}}^{P} \sum_{l=0}^{P} \alpha_{kl}\, x(i-k, j-l)\right]^2$$

is minimized

$$E\left[x(i,j) - \sum_{k=0}^{P} \sum_{l=0}^{P} \alpha_{kl}\, x(i-k, j-l)\right] x(m,n) = 0$$

$$\forall \quad m = i, i-1, \ldots, i-k$$
$$n = j, j-1, \ldots, j-l$$

Hence, $\alpha_{k\ell}$ are found by solving

$$A\underline{\alpha} = b$$

where elements of vector α are the coefficients $\alpha_{k\ell}$ and the elements of the matrix A and vector b are values of the correlation $R(m,n)$

$$\sigma^2 = E\left[x(i,j) - \sum_{\substack{k=0 \\ k+\ell\neq 0}}^{P}\sum_{\ell=0}^{P}\alpha_{k\ell}\, x(i-k, j-\ell)\right] x(i,j)$$

$$- \underbrace{E\left[x(i,j) - \sum\sum \alpha_{k\ell}\, x(i-k, j-\ell)\right]\left[\sum\sum x(i-k, j-\ell)\right]}_{0}$$

$$\sigma^2 = R(0,0) - \sum_{\substack{k=0 \\ k+\ell\neq 0}}^{P}\sum_{\ell=0}^{P}\alpha_{k\ell}\, E\, x(i,j)\, x(i-k, j-\ell)$$

Example:

$$R(\tau_1, \tau_2) = \sigma_s^2 \, e^{-a|\tau_1| - b|\tau_2|}$$

$$x(i,j) = \alpha_1 x(i, j-1) + \alpha_2 x(i-1, j) + \alpha_3 x(i-1, j-1) + \sigma u(i,j)$$

$$E\left[x(i,j) - \alpha_1 x(i, j-1) + \alpha_2 x(i-1,j) + \alpha_3 x(i-1,j-1)\right] \begin{matrix} x(i,j-1) \\ x(i-1,j) \\ x(i-1,j-1) \end{matrix} = 0$$

$$\begin{bmatrix} R_{00} & R_{11} & R_{10} \\ R_{11} & R_{00} & R_{01} \\ R_{10} & R_{01} & R_{00} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} R_{01} \\ R_{10} \\ R_{11} \end{bmatrix}$$

$$\alpha_1 = \frac{R_{01}}{\sigma_s^2} = e^{-b}$$

$$\alpha_2 = \frac{R_{10}}{\sigma_s^2} = e^{-a}$$

$$\alpha_3 = -\frac{R_{11}}{\sigma_s^2} = -e^{-a-b} = -\alpha_1 \alpha_2$$

$$\sigma^2 = \sigma_s^2 \left(1 - \alpha_1^2 - \alpha_2^2 + \alpha_3^2\right)$$

Choosing model's order & stability:

In general, successively higher order models are assumed and their coefficients a_k^p and σ_p^2 are computed. Optimal choice of P is made according to one or more of the following

1- σ^2 does not change with increasing P i.e.

$$\sigma_{p+1}^2 = \sigma_p^2$$

2- Only few values of the correlation function are available

3- Rate of decrease of σ_p^2 as P increases

4- trade off between the decrease in σ_p^2 and additional implementation complexity and cost.

STABILITY

Properties of the modelling procedure:

1- If the underlying 2-D process satisfies a finite order autoregressive model, Then this procedure will find that model

2- The correlation generated by the approxima model matches the apriori correlation at, at least, M points, where
M = total number of model coefficients

3- Only numerical values of R(m, n) are needed and no analytic form is required

4- Nonstationarity does not pose a grave problem

SEGMENTATION - CLASSIFICATION

OPTIMALITY:

For a set X

$$X = \{ x_1, x_2, \ldots\ldots, x_N \}$$

in a two class environment ω_1 and ω_2, the segmentation $X_1 \in \omega_1$ and $X_2 \in \omega_2$ is optimal iff

$$p(X_1 | \omega_1)\, p(X_2 | \omega_2) \geqslant p(Z_1 | \omega_1)\, p(Z_2 | \omega_2)$$

for any other segments Z_1 & Z_2

where

$$Z_1 \cap Z_2 = \phi$$

$$Z_1 \cup Z_2 = X$$

SEGMENTATION - USING AUTOREGRESSIVE MODELS

$$x(\cdot,\cdot) \in \omega_n$$

$$x(i,j) - \mu_n = \sum_{\substack{k=0 \\ k=l\neq 0}}^{P_n} \sum_{l=0}^{P_n} \alpha_{kl}^{n} \left[x(i-k, j-l) - \mu_n \right] + \sigma_n u(i,j)$$

In 1-D environment

$$x(k+1) - \mu_n = \sum_{i=0}^{P_n} \alpha_i^{n} \left[x(k-i) - \mu_n \right] + \sigma_n u(k)$$

$$X = \{ x_1, x_2, \ldots\ldots, x_N \}$$

$$p(x_1, \ldots\ldots, x_N \mid \omega_i)$$

$$p(x_1, \ldots, x_N / \omega_j)$$

1- USE OF MODELS AS WHITENING FILTERS DISCREDITTED

$$p(x_1, \ldots, x_N \mid \omega_n)$$

$$x(k) - \mu_n = \sum_{i=1}^{P_n} \alpha_i^n [x(k-i) - \mu_n] + u(k-1)$$

$$E\, u^2(j) = \sigma_n^2$$

$$u(k-1) = x(k) - \mu_n - \sum \alpha_i^n [x(k-i) - \mu_n]$$

$$p(x_1, \ldots, x_N \mid \omega_n)$$

$$p(u_1, \ldots, u_N \mid \omega_n)$$

$$= p(u_1 \mid \omega_n)\, p(u_2 \mid \omega_n) \ldots p(u_N \mid \omega_n)$$

so that optimal classification becomes a pixel by pixel operation such that

$x_j \in \omega_n$ if

$$p(u_j \mid \omega_n) \geq p(u_j \mid \omega_i) \quad \forall\, i \neq n$$

EXAMPLE 1 - DIFFERENT MEANS

① $$x(k+1) = 0.99\, x(k) + u_1(k)$$

② $$x(k+1) = 0.99\, x(k) + u_2(k)$$

$$\sigma_1^2 = E\, u_1^2(k) = \sigma_2^2 = E\, u_2^2(k) = 1$$

$$\mu_1 = 10 \qquad \mu_2 = 80$$

$$y(k) = 11 \qquad y(k+1) = 12$$

$$\hat{x}_1(k+1) = 0.99(11-10) = 0.99$$

$$\hat{x}_2(k+1) = 0.99(11-80) = -68.31$$

$$c_1 = \ln 1 + (12-10-0.99)^2 = 1.02$$

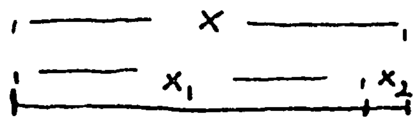
$$c_2 = \ln 1 + (12-80+68.31)^2 = 0.1$$

$$c_1' = \ln 1 + (12-10-1.98)^2 = 0.0004$$

$$c_2' = \ln 1 + (12-80+136.62)^2 = 4708.7$$

2 - A SEQUENTIAL SEGMENTATION METHOD

1-D EXAMPLE



Let

$$p(x_1 \mid \omega_1) > p(x_2 \mid \omega_2)$$

DECISION RULE:

$x_1 \in \omega_1$

COMPUTE:

$$p(x_2 \mid x_1, \omega_1) \quad \& \quad p(x_2 \mid \omega_2)$$

FROM CLASS 1 MODEL FROM APRIORI STATS

IF $p(x_2 \mid x_1, \omega_1) > p(x_2 \mid \omega_2)$ THEN $x_2 \in \omega_1$

IF $p(x_2 \mid \omega_2) > p(x_2 \mid x_1, \omega_1)$ THEN $x_2 \in \omega_2$

OPTIMALITY OF THE DECISION RULE

$$p(x_1|\omega_1) > p(x_1|\omega_2)$$

SUPPOSE

$$p(x_2|x_1,\omega_1) > p(x_2|\omega_2)$$

$$x_1, x_2 \in \omega_1$$

$$p(x_1|\omega_1)\,p(x_2|x_1,\omega_1) > p(x_1|\omega_1)\,p(x_2|\omega_2)$$

$$\underline{p(x_1,x_2|\omega_1) > p(x_1|\omega_1)\,p(x_2|\omega_2)}$$

AND IF

$$p(x_2|\omega_2) > p(x_2|x_1,\omega_1)$$

$$x_1 \in \omega_1 \qquad x_2 \in \omega_2$$

$$\underline{p(x_2|\omega_2)\,p(x_1|\omega_1) > p(x_1,x_2|\omega_1)}$$

$$p(x_1|\omega_1) > p(x_1|\omega_2)$$

$$p(x_2|x_1,\omega_1) > p(x_2|\omega_2)$$

$$x_1, x_2 \in \omega_1$$

$$p(x_1,x_2|\omega_1) > p(x_1|\omega_2)\, p(x_2|\omega_2)$$

$$\overset{?}{} \quad p(x_1,x_2|\omega_2)$$

$$p(x_1|\omega_2)\, p(x_2|\omega_1) < p(x_1|\omega_1)\, p(x_2|\omega_1)$$

$$\overset{?}{<} \quad p(x_1,x_2|\omega_1)$$

N83
23068

UNCLAS

This Page Intentionally Left Blank

FUN.STAT and Statistical Image Representations

Notes for Presentation by

Emanuel Parzen

Institute of Statistics, Texas A&M University

at

NASA/MPIRA Workshop (Math. Stat.)

January 27-28, 1983

Abstract

Presentation consists of: (1) outline of general ideas of functional statistical inference analysis of one sample and two samples, univariate and bivariate, and (2) application of ONESAM program to analyze the univariate probability distributions of multi-spectral image data.

MULTI-SPECTRAL IMAGE DATA ANALYZED BY ONESAM PROGRAM

Data analyzed consists of 9 files. Each file represents an observation at a different time in the growing season of a geographical area measuring 5 by 6 nautical miles divided into segments 117 across and 196 down, for a total of 22932 picture segments. The successive files represent flights on days

128, 145, 146, 163, 182, 199, 200, 235, 236.

There are 4 measurements per element, representing 4 channels recorded by a multi-spectral scanner of sunlight reflection. The channels are respectively 4, 5, 6, 7 angstroms, going from visible to infrared to far infrared bands of the spectrum. The first two channels may be similar (highly related), and the second two channels may be similar.

For this area, ground truth data is available in which each element is divided into 6 subpixels. The data was originally collected for LACIE (Large Area Crop Inventory Experiment).

There are 36 data sets representing 4 channels in 9 files. Each data set consists of 22392 integers (theoretically from 0 to 256) representing the reflections from surface elements. The data was received by us from Dr. Guseman in the form of 36 historgrams. The histograms were analyzed by our ONESAM program to determine the shape of the distribution fitting the histogram, and in particular to determine: (1) if the distribution is unimodal or bimodal; (2) the variation of medians and interquartile ranges.

This research aims to contribute to, among other problems, digital image representation whose definition we quote.

> Digital image representation is the determination and modeling of basic characteristics or features of the digital image which can be incorporated into the process of identifying classes and attributes in a scene. Approaches to the modeling of spatial image characteristics that require research include quantitative descriptions of image texture and the segmentation of images on the basis of spatial structure. Research is needed to determine the scene probability density functions and class conditional density functions of digital image data in order to understand spectral characteristics and extract desired information. Determination of density functions will enable the development of data transformation which reduce the dimensions of multi-variate image data while preserving information pertaining to scene classes and attributes.

Entropy difference estimator
using autoregressive order (m)

Day	File	Channel	Median	I.Q.	m=1	m=2	m=3	m=4	m=5	Best AIC Order
128	1	1	24.6	5.31	.092	.145	.153	.161	.178	2
145	2	1	27.5	3.85	.154	.179	.330	.386	.391	4
146	3	1	26.2	3.73	.056	.195	.204	.228	.236	2
163	4	1	27.4	4.94	.027	.061	.151	.154	.180	3
182	5	1	21.4	4.93	.236	.323	.365	.376	.391	3
199	6	1	22.8	4.40	.325	.402	.413	.426	.430	2
200	7	1	37.3	8.21	.063	.094	.094	.097	.106	1
235	8	1	32.0	13.9	.034	.048	.052	.055	.055	1
236	9	1	34.2	9.77	.025	.027	.027	.031	.044	1

Day	File	Channel	Median	I.Q.	m=1	m=2	m=3	m=4	m=5	Best AIC Order
128	1	2	24.9	7.05	.059	.144	.152	.203	.203	2
145	2	2	26.2	4.21	.125	.158	.182	.218	.223	1
146	3	2	25.0	4.27	.109	.184	.196	.206	.223	2
163	4	2	22.3	7.23	.064	.105	.103	.129	.146	2
182	5	2	15.2	9.03	.145	.289	.329	.355	.378	5
199	6	2	19.0	6.62	.153	.315	.321	.347	.348	2
200	7	2	31.9	9.16	.063	.070	.074	.086	.092	1
235	8	2	38.0	26.3	.059	.080	.080	.090	.092	1
236	9	2	39.5	20.9	.050	.064	.066	.073	.074	1

Day	File	Channel	Median	I.Q.	m=1	m=2	m=3	m=4	m=5	Best AIC Order
128	1	3	26.1	10.6	.029	.066	.085	.091	.102	1
145	2	3	30.8	8.64	.154	.182	.204	.206	.208	1
146	3	3	30.5	8.90	.118	.141	.149	.171	.187	1
163	4	3	44.3	14.7	.004	.014	.019	.026	.033	1*
182	5	3	55.0	19.2	.014	.016	.023	.029	.029	1*
199	6	3	54.8	15.9	.043	.050	.051	.054	.058	1
200	7	3	65.0	15.0	.013	.027	.033	.046	.053	1*
235	8	3	51.2	20.7	.001	.019	.022	.026	.031	1*
236	9	3	49.6	17.2	.002	.018	.021	.022	.025	1*

Day	File	Channel	Median	I.Q.	m=1	m=2	m=3	m=4	m=5	Best AIC Order
128	1	4	13.5	4.23	.033	.080	.097	.107	.110	1
145	2	4	15.0	3.95	.212	.253	.273	.279	.282	1
146	3	4	14.6	4.02	.165	.201	.223	.231	.235	1
163	4	4	20.4	7.54	.004	.021	.029	.041	.044	1*
182	5	4	28.2	11.9	.005	.007	.022	.034	.039	1*
199	6	4	27.8	8.40	.043	.049	.052	.052	.053	1
200	7	4	29.7	7.10	.001	.013	.016	.018	.019	1*
235	8	4	23.6	7.85	.006	.031	.032	.033	.035	1*
236	9	4	22.1	6.76	.008	.022	.027	.033	.037	1*

Handout for remarks by Professor Emanuel Parzen

FUNCTIONAL STATISTICAL INFERENCE

FUN.STAT APPROACH TO DENSITY ESTIMATION

1. ONE SAMPLE: UNIVARIATE

Let X be continuous random variable and $X_1, X_2, \ldots, X_n$ a random sample of X. To estimate distribution function

$$F_X(x) = \Pr[X \leq x]$$

and probability density $f(x) = F'(x)$, we estimate quantile function

$$Q_X(u) = F_X^{-1}(u),$$

quantile density $q_X(u) = Q_X'(u)$, and density quantile

$$fQ_X(u) = f_X(Q_X(u)).$$

1. Form sample distribution function $\tilde{F}_X(x)$, sample quantile function $\tilde{Q}_X(u)$, sample quantile density $\tilde{q}(u)$ at $u = j/(n+1)$, $j=1,2,\ldots,n$.
2. Plot sample version of informative quantile function

$$IQ(u) = \frac{Q(u)-Q(0.5)}{2\{Q(0.75) - Q(0.25)\}}$$

whose values as u tends to 0 and 1 indicates the tail exponents of the probability law of X.

3. Determine a standard distribution function $F_o(x)$ to test

$$H_o: F(x) = F_o\left(\frac{x-\mu}{\sigma}\right) \text{ or } Q(u) = \mu + \sigma\, Q_o(u)$$

for location and scale parameters μ and σ to be estimated. Form

$$\tilde{d}(u) = f_oQ_o(u)\ \tilde{q}(u) \div \tilde{\sigma}_o \quad ,$$

$$\tilde{\sigma}_o = \int_o^1 f_oQ_o(t)\ \tilde{q}(t)\ dt.$$

which estimate respectively

$$d(u) = f_oQ_o(u)\ q(u) \div \sigma_o \quad ,$$

$$\sigma_o = \int_o^1 f_oQ_o(t)\ q(t)\ dt.$$

4. Form successive autoregressive estimators

$$\hat{d}_m(u) = \hat{K}_m \left| 1 + \hat{\alpha}_m(1)\ e^{2\pi iu} + \ldots + \hat{\alpha}_m(m)\ e^{2\pi ium} \right|^{-2}$$

whose negentropy

$$\hat{H}_m = \int_o^1 - \log \hat{d}_m(u)\ du = -\log \hat{K}_m$$

is used to determine optimal orders $\hat{m}$. Note that $\hat{H}_m$ estimates the entropy difference

$$\Delta = \{\log \sigma_o - \int_0^1 \log f_oQ_o(u)\} - \{- \int_0^1 \log fQ(u)du\}$$

5. Estimate $fQ(u)$ by

$$\hat{f}Q_m(u) = f_oQ_o(u) \div \bar{\sigma}_o \hat{d}_m(u)$$

where m is chosen equal to an optimal order $\hat{m}$.

2. TWO SAMPLE: UNIVARIATE

Let X and Y be continuous random variables with random samples $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$ respectively, and with respective distribution functions

$$F(x) = Pr[X\leq x], \qquad G(x) = Pr[Y\leq x].$$

The pooled sample $X_1,\ldots,X_m$, $Y_1,\ldots,Y_n$ can be regarded as a random sample from the distribution function

$$H(x) = \lambda\ F(x) + (1-\lambda)\ G(x), \quad \lambda = \frac{m}{m+n} \quad .$$

To test the hypotheses of equality of distributions,

$$H_o:\ F(x) = G(x) = H(x),$$

it is customary in non-parametric statistics to introduce

$$D_X(u) = F\,H^{-1}(u), \qquad D_Y(u) = G\,H^{-1}(u)$$

with densities

$$d_X(u) = \frac{f\,H^{-1}(u)}{h\,H^{-1}(u)} \quad , \qquad d_Y(u) = \frac{g\,H^{-1}(u)}{h\,H^{-1}(u)}$$

Note that $h\,H^{-1}(u) = \lambda\,f\,H^{-1}(u) + (i-\ )\,g\,H^{-1}(u)$; therefore

$$d_X(u) = \left\{\lambda + (1-\lambda)\,\frac{g\,H^{-1}(u)}{f\,H^{-1}(u)}\right\}^{-1}$$

A raw estimator of $D_X(u)$ is

$$\tilde{D}_X(u) = \tilde{F}\,\tilde{H}^{-1}(u)$$

from which one can form

$$\tilde{\rho}(v) = \int_0^1 e^{2\pi i u v}\, d\,\tilde{D}_X(u)$$

and autoregressive estimators $\hat{d}_{X,m}(u)$ of $d_X(u)$.

When one observes k variables $X^{(1)}, X^{(2)}, \ldots, X^{(k)}$, one estimates (for $j=1,\ldots,k$) the densities of $D_j(u) = F_{X(j)}\,(H^{-1}(u))$.

3. ONE SAMPLE: BIVARIATE

Let (X_1, X_2) be jointly continuous random variables with distribution function

$$F_{X_1,X_2}(x_1, x_2) = Pr[X_1 \le x_1, X_2 \le x_2]$$

and density $f_{X_1,X_2}(x_1, x_2)$. The joint density quantile function is defined by

$$fQ_{X_1,X_2}(u_1,u_2) = f_{X_1,X_2}(Q_{X_1}(u_1), Q_{X_2}(u_2))$$

To estimate fQ we define

$$D_{X_1,X_2}(u_1,u_2) = F_{X_1,X_2}(Q_{X_1}(u_1), Q_{X_2}(u_2))$$

which is the distribution function of $U_1 = F_{X_1}(X_1)$, $U_2 = F_{X_2}(X_2)$; it has density

$$d_{X_1,X_2}(u_1,u_2) = \frac{\partial^2}{\partial u_1, \partial u_2} D(u_1,u_2)$$

satisfying

$$fQ_{X_1,X_2}(u_1,u_2) = fQ_{X_1}(u_1)\, fQ_{X_2}(u_2)\, d_{X_1,X_2}(u_1,u_2).$$

To estimate d_{X_1,X_2} from a random sample $(X_1^{(j)},X_2^{(j)})$, j=1,...,n, form

$$\tilde{D}_{X_1,X_2} = \tilde{F}_{X_1,X_2}(\tilde{Q}_{X_1}(u_1), \tilde{Q}_{X_2}(u_2))$$

and a raw estimator $\tilde{d}_{X_1,X_2}(u_1,u_2)$. We smooth $\log \tilde{d}_{X_1,X_2}(u_1,u_2)$ by a smooth estimator $\log \hat{d}_{X_1,X_2}(u_1,u_2)$ minimizing a criterion similar to

$$\sum_{j=1}^{n} | \log \tilde{d}[U_1^{(j)},U_2^{(j)}] - \log d_m[U_1^{(j)},U_2^{(j)}]|^2$$

where $\log d_m(u_1,u_2)$ has the parametric representation

$$\log d_m(u_1,u_2) = \sum_{\nu_1,\nu_2} \theta_{\nu_1,\nu_2} \exp i\ (u_1\nu_1 + u_2\nu_2) - \psi(\theta_{\nu_1,\nu_2}) \quad ;$$

where the summation is over $\nu_1,\nu_2 = 0, \pm 1,\ldots,\pm m$,

and $\psi(\theta_{\nu_1,\nu_2})$ is an integrating factor to make $d_m(u_1,u_2)$ a probability density. The foregoing estimators have been implemented in the Ph.D. Thesis of T. J. Woodfield. The problem of choosing a best value of the order m is approached by evaluating the entropy of d_m.

We expect Woodfield to work with us this summer to extend his results to estimation of multivariate density quantile functions.

4. TWO SAMPLES: BIVARIATE

Let (X_1,X_2) and (Y_1,Y_2) be random vectors with respective distribution functions $F(X_1,X_2)$ and $G(X_1,Y_2)$, and respective random samples

$$(X_1^{(j)},X_2^{(j)}),\ j=1,\ldots,m \quad \text{and} \quad (Y_1^{(k)},Y_2^{(k)}),\ k=1,2,\ldots,n.$$

Let $H(x_1,x_2)$ denote the distribution function of the pooled random sample, with marginal distribution functions $H_1(x_1)$ and $H_2(x_2)$. Define

$$D_1(u_1,u_2) = F(H_1^{-1}(u_1),\ H_2^{-1}(u)) \quad ,$$

$$D_2(u_1,u_2) = G(H_1^{-1}(u_1),\ H_2^{-1}(u_2))$$

From $D_1(u_1,u_2)$ and $D_2(u_1,u_2)$ one can form raw estimators $d_1(u_1,u_2)$ and $d_2(u_1,u_2)$ of the densities

$$d_1(u_1,u_2) = \frac{f(H_1^{-1}(u_1),\ H_2^{-1}(u_2))}{h_1H_1^{-1}(u_1)\ h_2H_2^{-1}(u_2)} \quad ,$$

$$d_2(u_1,u_2) = \frac{g(H_1^{-1}(u_1),\ H_2^{-1}(u_2))}{h_1H_1^{-1}(u_1)\ h_2H_2^{-1}(u)}$$

Therefore

$$\log d_1(u_1,u_2) - \log d_2(u_1,u_2)$$

$$= \log f(H_1^{-1}(u_1),\ H_2^{-1}(u_2)) - \log g(H_1^{-1}(u_1),\ H_2^{-1}(u_2))$$

The likelihood ratio $f(x_1,x_2)/g(x_1,x_2)$ can be effectively estimated by estimating $\log d_1(u_1,u_2) - \log d_2(u_1,u_2)$. We propose to investigate exponential model representations of

$$\log d_1(u_1,u_2) - \log d_{11}(u_1) - \log d_{12}(u_2)$$

where $d_{11}(u_1)$ and $d_{12}(u_2)$ are the marginal densities of $d_1(u_1,u_2)$ whic can be estimated by methods of two samples: univariate.

The final output are contour plots of the classification stacistic

$$L(x_1,x_2) = \log f(x_1,x_2) - \log g(x_1,x_2) \quad .$$

A point (x_1,x_2) is classified in population 1 or 2 by whether $L(x_1,x_2)$ exceeds a threshold which depends on the prior probabilities and loss funct'on.

This Page Intentionally Left Blank

ORIGINAL PAGE IS
OF POOR QUALITY

Omit to p. 101

77

STATISTICAL IMAGE REPRESENTATIONS:
Non-Gaussian Classification

Presentation Outline by

William B. Smith
Institute of Statistics, Texas A&M University

at

NASA/MPIRA Workshop (Math. Stat.)
January 27-28, 1983

I. Introduction

II. Two Population Problem

Populations: Normal vs. Normal
Mixed Normal vs. Mixed Normal
Non-normal vs. Non-normal

Classifiers: Bayes
LDF
QDF

Measures of Non-normality:
Ashikaga - N*
Malkovich - Afifi - skewness/kurtosis
Mardia - skewness/kurtosis
Misclassification probability

Performance of standard tests -
Hotelling's T^2 -
Max. characteristic root -

III. Multipopulation problems
Canonical Correlation - several LDF
Missing data -
Utilization decision for partial records - AIC

IV. Summary.

PRECEDING PAGE BLANK NOT FILMED

PAGE 76 INTENTIONALLY BLANK

Univariate Data--2 populations

Normal vs. Normal

		E(Y)	Var(Y)	Δ	$\Phi(-\Delta/2)$
Pops	π_1	0	1		
	π_2	1	1	1	.31
		2	1	2	.16
		3	1	3	.07

Lognormal vs. Lognormal

		E(X)	Var(X)	$\sqrt{\beta_1}$	β_2	Δ^*
Pops	π_1	1.65	4.67	6.2	113.9	
	π_2	4.48	34.51	↓	↓	.64
		12.18	255.02			.92
		33.12	1884.32			1.02

		L_1	L_2	$\frac{1}{2}(L_1+L_2)$	Q_1	Q_2	$\frac{Q_1+Q_2}{2}$
Pops	π_1	.55	.13	.34	.79	.04	.42
	π_2	.47	.03	.25	.73	.005	.37
		.44	.002	.22	.22	.013	.11

$$\Delta^* = \left[\left(E(x|\pi_1) - E(x|\pi_2)\right)^2 / \sigma_p^2\right]^{1/2} \qquad \sigma_p^2 = \frac{\sigma_1^2+\sigma_2^2}{2}$$

L_i = P($x \in \pi_i$ misclass using LDF)

Q_i = P($x \in \pi_i$ " " QDF).

Slide 1 revised

Untransformed data

	E(y)	Var(y)	Δ^*	$\frac{Q_1+Q_2}{2}$
π_1	0	1		
π_2	.848	.4394	1	.30
	1.530	.1698	2	.13
	2.179	.0551	3	.04

Transformed data

	E(x)	Var(y)	Δ	$\sqrt{\beta_1}$	β_2
π_1	1.65	4.67		6.2	113.
π_2	2.91	4.67	.58	2.6	17.
	5.03	4.67	1.56	1.4	6.
	9.08	4.67	3.44	0.7	4.

Transformed data

	L_1	L_2	$\frac{L_1+L_2}{2}$	$\Phi(-\Delta/$
π_1				
π_2	.21	.48	.35	.39
	.11	.21	.17	.22
	.05	.02	.035	.048

Discrimination of k (>2) normal populations

Let $\pi_i : N_p(\mu_i, \Sigma)$ $i = 1, \ldots k$.

Define E = within sums of squares and products matrix

H = between population sums of squares and products matrix

Standard tests of the identity of the k pops involve HE^{-1} — or equivalently finding a vector ℓ to maximize

$$\frac{\ell' H \ell}{\ell' (H+E) \ell}.$$

It is easily shown that the "optimum" ℓ must satisfy

$$\left[-r^2(H+E) + H\right] \ell = 0$$

where r satisfies

$$\det\left[-r^2(H+E) + H\right] = 0$$

If the observations X are each augmented by a dummy vector y which identifies their parent population — ie

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_g \end{pmatrix} \qquad y_j = \begin{cases} 0, & x_i \notin \pi_j \\ 1, & x_i \in \pi_j, \quad g = k-1 \end{cases}$$

Then the equation

$$\det\left(-r^2 (H+E) + E\right) = 0$$

has $f = \min(p, g)$ roots $r_1^2 \geq r_2^2 \cdots \geq r_f^2$.

These r_i's are the squared canonical correlations between x & y and their corresponding canonical vectors (say, l_i) determine k discriminant functions

$$l_i^t x .$$

Canonical Correlations

Given $Var\begin{pmatrix} \underset{\sim}{x} \\ \cdots \\ \underset{\sim}{y} \end{pmatrix} = \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ & \Sigma_{22} \end{bmatrix}$

then the canonical correlation coefficients relating $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are the r solutions to

$$\left| -\rho^2 \Sigma_{11} + \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right| = 0 \qquad (*)$$

When estimating $\rho_1, \cdots, \rho_r$ when only full data vectors are observed, one replaces Σ_{ij} with the appropriate estimates.

When both full and partial records are observed, we proposed two techniques

1. Estimate ρ_i's directly using H-S
2. Estimate Σ using HM and substitute into $(*)$

Simulation Study

p = dim. of $\underset{\sim}{X}$

q = dim. of $\underset{\sim}{Y}$

$$\Sigma = \begin{bmatrix} 1 & a & & a \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

$a = .4, .5, \ldots .9$ $\qquad N = 50$

n_1 = no. complete records

n_2 = no. partial records.

Note: $p = q = 1 \Rightarrow \rho^2 = a^2$

$p = 1 \quad q = 2 \Rightarrow \rho^2 = 2a^2/(1+a)$

$p = 2 \quad q = 2 \Rightarrow \rho_1^2 = 4a^2/(1+a)^2 \quad \rho_2^2 = 0.$

Table 6

Simulation of Example 3.3
$p = 1,\ q = 2$
$n_1 = 10,\ n_2 = 40$

Population		Full Data $\tilde{r}^2$		All Data Hocking-Smith		All Data Hocking-Marx	
a	ρ^2	Est. Bias	MSE	Est. Bias	MSE	Est. Bias	MSE
.4	.22857	.13567	.06176	.04640	.06338	.10519	.05072
.5	.33333	.10479	.05751	.03036	.06512	.07875	.03412
.6	.45000	.07241	.04747	.04040	.04407	.06480	.02698
.7	.57647	.03946	.03682	.04230	.03076	.04723	.02038
.8	.71111	.02414	.02437	.03325	.02425	.03154	.01044
.9	.85263	.00487	.00923	.02800	.01027	.01470	.00297

Table 7

Simulation with Example 3.4

$p = 2, q = 2$

$n_1 = 40, n_2 = 10$

Population		Full Data			All Data Hocking-Marx		
a	ρ_1^2	Est. Bias	Det. MSEx10^3	Trace	Est. Bias	Det. MSEx10^3	Trace
.4	.32653	.04883	.02471	.01650	.04479	.01906	.01483
.5	.44444	.04347	.02592	.01564	.04116	.02160	.01415
.6	.56250	.03463	.01757	.01216	.03301	.01267	.01046
.7	.67820	.02898	.01202	.00846	.02678	.01032	.00776
.8	.79012	.02811	.00787	.00583	.02691	.00567	.00503
.9	.89751	.02764	.00196	.00304	.02598	.00157	.00270

Table 8

Simulation with Example 3.4

$p = 2,\ q = 2$

$n_1 = 20,\ n_2 = 30$

Population		Full Data			All Data Hocking- Marx		
a	ρ_1^2	Est. Bias	Det. MSE$\times 10^3$	Trace	Est. Bias	Det. MSE$\times 10^3$	Trace
.4	.32653	.10331	.16013	.03897	.08391	.09417	.02835
.5	.44444	.07997	.15578	.03391	.06511	.08815	.02444
.6	.56250	.06436	.11410	.02544	.05484	.06114	.01779
.7	.67820	.05895	.08405	.02067	.05020	.04097	.01373
.8	.79012	.05327	.04995	.01428	.04692	.02382	.00988
.9	.89751	.05384	.01465	.00963	.04662	.00659	.00697

Akaike Information

$$AIC = -2 \ln L + 2k$$

k = # parameters (Akaike 1974, 1979)

Derive test for significance of partial records taking the difference between AIC evaluated using "full" data estimates only and AIC evaluated at the combined (HM) estimates.

Example III.

n_1 obs from $N_p(\mu, \Sigma)$, Σ known

n_2 obs from $N_q(\mu_2, \Sigma_2)$, $q<p$

$\mu_2 = D\mu$, $\Sigma_2 = D\Sigma D'$

Then $\tilde{\mu} = \hat{\mu}_1 + B(D\hat{\mu}_1 - \hat{\mu}_2)$

$$B = -\frac{n_2}{N} \Sigma D' (D\Sigma D')^{-1}$$

So $Q_1 = AIC(\sim) - AIC(\wedge)$

$$= -\frac{n_2}{n_1} \chi^2_{calc}$$

where 88 $$\chi^2_{calc} = \frac{n_1 n_2}{N}\left(\hat{\mu}_2 - D\hat{\mu}_1\right)'(D\Sigma D')^{-1}\left(\hat{\mu}_2 - D\hat{\mu}_1\right)$$

$$\sim \chi^2_q$$

Note: Q_1 is a function of the Mahalanobis distance

Q_1 is a form of Hotelling's T^2 (variance known).

<u>Example IV</u>: Assume Σ unknown in Example <u>III</u>.

Since $\tilde{\mu}$, $\tilde{\Sigma}$ are m.l.e. for μ & Σ, we have

$$Q_2 = AIC(\sim) - AIC(\wedge) \leq 0.$$

In fact, Q_2 can be expressed as a decreasing function of

$$T_1^2 = Z'(D\hat{\Sigma}_1 D')^{-1} Z$$

and $$T_2^2 = tr\left[(D\hat{\Sigma}_1 D')^{-1}\hat{\Sigma}_2\right]$$

where $$Z = \hat{\mu}_2 - D\hat{\mu}_1 .$$

$Z = \hat{\mu}_2 - D\hat{\mu}_1$

Example III – detail

$$AIC_1 = -(\text{const}) + n_1 \ln|\Sigma| + n_2 \ln|\Sigma_2|$$
$$+ n_1 \operatorname{tr} \Sigma^{-1}\hat{\Sigma}_1 + n_2 \operatorname{tr} \Sigma_2^{-1}\left[\hat{\Sigma}_2 + (\hat{\mu}_2 - D\hat{\mu}_1)(\hat{\mu}_2 - D\hat{\mu}\right.$$
$$+ 2p$$

$$AIC_{12} = -(\text{const}) + n_1 \ln|\Sigma| + n_2 \ln|\Sigma_2|$$
$$+ n_2 \operatorname{tr} \Sigma^{-1}\hat{\Sigma}_1 + n_2 \operatorname{tr} \Sigma_2^{-1}\left[\hat{\Sigma}_2 + (\hat{\mu}_2 - D\tilde{\mu})(\hat{\mu}_2 - D\right.$$
$$+ 2p$$

Recommendations:

1. Test the "significance" of the partial records available.

2. When doing canonical correlation with partial records, estimate the covariance matrix (using HM, say). Then input the $\tilde{\Sigma}$ matrix as a

   TYPE = COV

   matrix in PROC CANCORR or input $\tilde{\Sigma}$ into PROC MATRIX and extract the appropriate eigenvalues and eigenvectors.

ORIGINAL PAGE IS
OF POOR QUALITY

Multivariate Time Series in Two Dimensions and the Classification Problem

Outline of Presentation

H.J. Newton

NASA/MPIRA Workshop
January 27-28, 1983

I. Introduction and Basic Aims of Research

II. Transects and Classification of Model Signatures

III. An Analysis of Variance Approach to Finding Boundaries

IV. Incorporating Temporal Correlation

V. Incorporating the Second Dimension

VI. Some Computational Considerations

I. Introduction And Basic Aims

$\underset{\sim}{Y}(j,k;t)$ = 4 dimensional random vector at time t at spatial index (j,k)

∴ Multivariate (4 dimensional) time series in 2 dimensions.

__Question__ How can spatial and temporal correlation be used in a procedure for:

1) Classifying a given scene as being heterogeneous or homogeneous
2) If homogeneous, what does it contain
3) If heterogeneous, are there discernible boundaries between homogeneous subscenes
4) Reducing the amount of data required to automate classification.

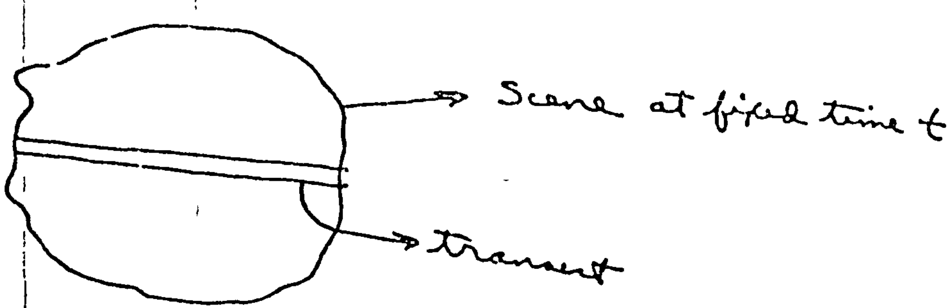
Suggested Answer Develop time series models having certain "signatures" that will vary more between scene types than within a scene type.

Basic Difficulties in Implementation

1) Theoretical : little is known of models in the general setting. Certainly little of estimation of parameters.

2) Algorithmic : The obvious extensions of the theory in simple cases to the general case require algorithms that are impractical.

II. Transects and Classification by Model Signatures



Let $\underset{\sim}{Y}(\ell)$ = 4-dimensional random vector at position ℓ in the transect.

$$K(\ell, m) = \operatorname{Cov}\left(\underset{\sim}{Y}(\ell), \underset{\sim}{Y}(m)\right)$$

Homogeneous $\Rightarrow$

$$K(\ell, m) = R(|\ell - m|)$$

ORIGINAL PAGE IS
OF POOR QUALITY

$$f(\omega) = \sum_{v=-\infty}^{\infty} R(v)\, e^{-2\pi i v \omega} \qquad , \omega \in [-1/2, 1/2]$$

$$\underset{\sim}{Y}(\ell) + \sum_{j=1}^{p} A(j)\, \underset{\sim}{Y}(\ell - j) = \underset{\sim}{\epsilon}(t)$$

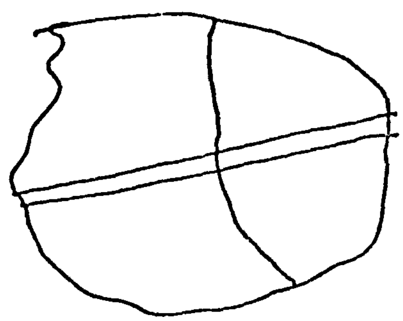
Estimate R, f, p, A and use for classification.

These estimators should behave differently for

1) homogeneous vs heterogeneous scenes

2) type i vs type j homogeneous scenes

Boundaries



Recursive estimation with change detection capability.

III. An ANOVA Approach to Finding Boundaries

- Let $\underset{\sim}{m}(\ell) = E(\underset{\sim}{Y}(\ell))$

- Do Several Transects

- Each transect can be considered as a sample from multivariate distribution having correlation between observations.

- Thus one has analysis of variance type problems with correlation (currently getting results for the analogous situation for univariate data over time).

- Heterogeneity and boundaries indicated by an ANOVA interaction between transect number and position in transect.

IV Temporal Correlation

May be difficult to discriminate between types at a fixed time but easily discriminate according to behavior over a period of time.

Thus:

V. Incorporating the Second Dimension

For fixed time, the transects are just multivariate time series with position in transect acting as time. To incorporate the second spatial dimension one can either:

1) Devise methods of combining transects (as in section III).

2) Devise methods that use models for explicitly relating $\underset{\sim}{Y}(j,k;t)$ to other $\underline{Y}$'s nearby.

Extension of R, S, ρ, A to two dimensions

Homogeneous: equally distant $\underset{\sim}{Y}$'s equally correlated

Traditional: $R(u,v) = \text{Cov}(\underset{\sim}{Y}(j,k), Y(j+u,k+v))$

$$f(\omega_1,\omega_2) = \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} R(u,v)\, e^{-2\pi i u \omega_1} e^{-2\pi i v \omega_2}$$

AR: $\underline{Y}(l,m)$ a linear function of $\underline{Y}$'s around it plus white noise error.

Currently working on what this means and the properties of various ways of defining "linear functions of $\underline{Y}$'s around $\underline{Y}(l,m)$".

VI. Some Computational Problems

Will need to develop fast algorithms for carrying out the procedures suggested by the above considerations.

In particular:

1) The theoretical structure of correlation matrices.

2) Fast recursive algorithms for estimation (such as Kalman Filter Algorithm)

N83
23069

UNCLAS

THE UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TEXAS 78712

A Minimax Approach to Spatial Estimation Using Affinity Matrices

Carl N. Morris

For presentation at NASA Project Meeting
Texas A&M, January 27-28, 1983

"An Empirical Bayes Approach to Some Spatial Analysis Problems, with Special Attention to Remotely Sensed Satellite Imagery".

Summary:

Our problem is to combine estimates made in the plane to improve on noisy unbiased estimates. We will want to use only a small fraction of points in a giant grid to do this, those that are most like a given point. Section 1 below provides a helpful component of this process defining an "affinity matrix" of values, indicating which points are relevant to others. Then Section 2 shows that minimax rules can be based on affinity matrices.

1. Affinity Matrices:

Let $a_{ij} \geq 0$ be the affinity of i for j, $1 \leq i, j \leq k$. We will assume for now that $A^{k \times k} = (a_{ij})$, the "affinity matrix", is symmetric and the rows (and columns) of A sum to unity. Someday we may wish to consider more general notions. In NASA applications, a_{ij} might be some diminishing function of d_{ij}, the distance from the i^{th} to the j^{th} pixel, but there could be cases where affinities are greater for more distant areas, e.g. crops growing in two valleys may be more similar to one another than to those growing on a mountain between them.

An affinity matrix is just a doubly stochastic matrix, one representing a reversible Markov chain, so much is known of them. For example, the eigenvalues $\alpha_1 \geq \ldots \geq \alpha_k$ satisfy $\alpha_1 = 1$ ($e = (1, \ldots, 1)'$ is its eigenvector), $\alpha_2 < 1$ (if the chain is irreducible), $\alpha_k \geq -1$. I don't know under what conditions $A \geq 0$, i.e. $\alpha_k \geq 0$, but we need to know. It may be that $A \geq 0$ if, under appropriate re-labeling, the elements a_{ij} diminish as they recede from the diagonal. Is $A \geq 0$, for example, if a_{ij} depends only on $|i-j|$ and decreases as $|i-j|$ increases?

Note that we can write

$$A = \sum_{i}^{k} \alpha_i P_i \tag{1.1}$$

with $P_1 = ee'/k$ and the $\{P_i\}$ a complete ($\Sigma P_i = I$) set of orthogonal projections, each with unit rank.

It's interesting to consider what affinities might be assigned to the center of a kxk grid k=5). Squared distances from the center are listed below.

8	5	4	5	8
5	2	1	2	5
4	1	0	1	4
5	2	1	2	5
8	5	4	5	8

#Pts.	Sq Dist
1	0
5	1
9	2
13	4
21	5
25	8

Fig. 1: Distances from center in a 5x5 grid.

This chart suggests nine and 21 point grids might be good. (Note: points outside the 5x5 grid all have squared distance greater than 8.) For 9 points, the a_{ij} might be: $a_{ii} = \frac{1}{4}$ (once), 1/8 (4 times), and 1/16 (4 times), which is given by

$$a_{ij} = 2^{-d_{ij}^2-2} \tag{1.2}$$

The point is that the two dimensional grid in Fig. 1 is collapsed into <u>one</u> row of an affinity matrix, with 9 non zero entries in the case of (1.2). This means we can ignore the complicated spatial structure in Fig. 2 when working out theories of estimation, at least in the <u>independence case</u> considered in the next section.

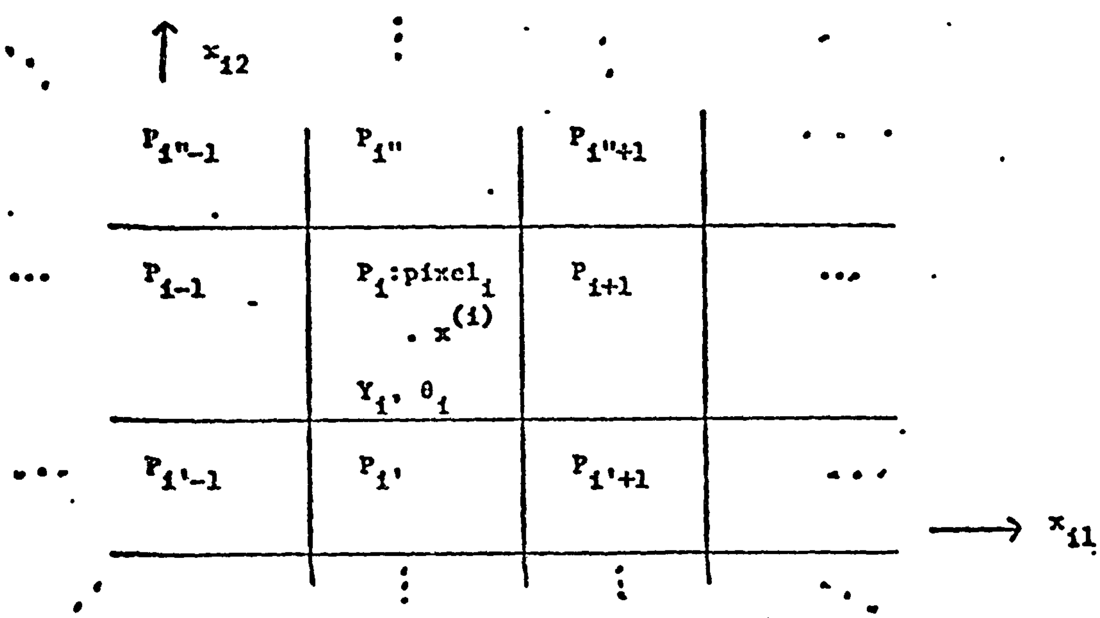


Figure 2. Areal problem organized into pixels or pixel-groups. Responses Y_i, true values θ_i.

2. A minimax rule for spatial analysis.

We take the simplest case of normal, independent observations (given the parameters) with equal (known) variances:

$$Y_i | \theta_i \overset{\text{ind}}{\sim} N(\theta_i, V), \; i=1,\dots, k.$$

We expect the $\{\theta_i\}$ to be correlated in any NASA application, but not necessarily the pixel intensity measurements $\{Y_i\}$ (there was some question raised about this at NASA last August). In practice the variances would have to be estimated, and perhaps this means we would actually be working with clusters of pixels. We need data to pursue these points, as well as the equality of variances and all other assumptions. Until such are obtained, we may proceed as follows.

Let A be any given kxk affinity matrix with non-negative eigenvalues $\alpha_1 = 1 > \alpha_2 \geq \dots \geq \alpha_k \geq 0$. Define

(2.1) $$Y^* = AY,$$

each Y_i^* being a weighted average of the $\{Y_j\}$. Let

(2.2) $$\hat{B} = \frac{(k-r-2)V}{\Sigma(Y_i - Y_i^*)^2}$$

with $r \equiv tr(A)$. Then let

(2.3) $$\hat{\theta}_i = (1-\hat{B})Y_i + \hat{B}\, Y_i^*$$

be the estimate of θ_i,

In case $A = ee'/k$, (2.1)-(2.3) reduce to Stein's estimator for shrinking all Y_i to $Y_i^* = \overline{Y}$. More generally, this estimator is that for shrinking toward a regression surface $E\theta = Z\beta$, Z^{kxr}, if $A = Z(Z'Z)^{-1}Z$ is an affinity matrix (since it needn't be, we can see that the affinity matrix assumption is unnecessary).

We now prove that the estimator (2.3) is minimax for loss

$$L(\theta,\hat{\theta}) = \Sigma(\hat{\theta}_i - \theta_i)^2/V \tag{2.4}$$

in the frequentist sense with risk $R(\theta)$ depending on θ. Stein's derivative formula

$$E_\theta(Y_i-\theta_i)\ f(Y_i) = V\ E_\theta f'(Y_i) \tag{2.5}$$

will be used where needed.

The following Lemma will be useful.

Lemma 1. Let M and T be symmetric matrices, t_i the i^{th} column vector of T, and $Q = Y'MY$. Then

$$\Sigma \frac{\partial}{\partial Y_i} \{\frac{1}{Q} t_i'Y\} = [Qtr(T)-2Y'MTY]/Q^2. \tag{2.7}$$

Theorem 1. The risk $R(\theta)$ of (2.3) has unbiased estimate

$$\hat{R} = k - (k-r-2)\hat{B} + 4\hat{B}(F-1) \tag{2.8}$$

with $F \equiv Y'(I-A)^3Y/Y'(I-A)^2Y$.

Thus, with α_k the minimum eigenvalue of A,

(2.9) $$\hat{R} \leq k - (k-r-2)\hat{B} - 4\hat{B}\alpha_k \text{ always}$$
$$\leq k - (k-r-2)\hat{B} \text{ if } \alpha_k \geq 0.$$

It follows that (2.3) is minimax if $k \geq r + 2$ when $\alpha_k \geq 0$.

Proof: Let $Q \equiv \Sigma(Y_i - Y_i^*)^2 = Y'(I-A)^2 Y$. The risk of (2.3) is $R(\theta) = E\,\|Y-\theta-\hat{B}(I-A)Y\|^2/V$

$$= k + E\hat{B}^2 Q/V - 2E\Sigma(Y_i-\theta_i)\,\hat{B}\,(I-A)_{(i)}Y/V$$

where $(I-A)_{(i)}$ is the i^{th} row vector of I-A. Now apply (2.5) to the last term and then (2.7) with $\hat{B} = cV/Q$, $c \equiv k-r-2$, and remove expectations to get

$$\hat{R} = k + c\hat{B} - 2\hat{B}\left[k-r-2\ \frac{Y'(I-A)^3Y}{Q}\right]$$
$$= k - \hat{B}[k-r-2\quad -4(F-1)]$$

Note that (2.9) follows because $0 \leq F \leq 1 - \alpha_k$. QED.

Of course, we always have

(2.10) $$\hat{R} < k - (k-r-6)\hat{B}$$

in (2.8) because $\alpha_k > -1$ for every affinity matrix.

We should be able to extend this proof to cover the case with $\hat{B}$ replaced in (2.3) by

(2.11) $$B^+ \equiv \min(\hat{B}, 1).$$

Further work must consider the following:

1. Let A be written as in (1.1) and write $S_i = Y'P_iY$.

Then the $\{S_i\}$ are independent non-central chi squares with one degree of freedom and expectation

$$ES_i = E\,\mathrm{tr}[P_i Y Y']$$
$$= V + \theta' P_i \theta,$$

ie.

$$S_i \overset{\text{ind}}{\sim} V\chi_1^{2\prime}\left(\frac{\theta' P_i \theta}{2V}\right).$$

Note that $\hat{B} = (k-r-2)V/\sum_i^k (1-\alpha_i)^2 S_i$ These facts may be useful in further development of the sampling properties of rules.

2. The rule will do well if $\theta'(I-A)^2\theta$ is fairly small. We will have to see if this is likely. Again, we need real data!

3. Is this nearly an empirical Bayes rule? For what (correlated) prior on the θ_i? What correlated priors, ones like those we might expect in NASA applications, lead to good empirical Bayes rules? With a one-dimensional spatial (e.g. time) problem, autoregressive priors seem to lead to estimators like Y*.

4. We need to find out a lot more theoretically about affinity matrices and their eigenstructure, and to consider which ones would be good for our applications. There will be problems near boundaries. Markov chain sources will be a good place to start. Covariance matrices arising in autoregressive theory also may be a useful source.

5. Note that the rule presented (2.3) is easy to compute. Still, we should consider carefully the computational aspects of this and any other rules we choose to derive.

6. Real applications will involve multivariate Y_i, say bivariate with Y_{i1} = greenness, Y_{i2} = brightness.

References:

Stein (1981), Annals of Statistics, pp. 1135-1151 has a result implying Theorem 1 here.

N83
23070

UNCLAS

THE UNIVERSITY OF TEXAS AT AUS
AUSTIN, TEXAS 78712

Department of Mathematics
RLM 8.100, RLM 10.130
(512) 471-7711

Hubert Kostal

Localized Shrinkage Factors and Minimax Results

Summary: A condition is derived under which a localized shrinkage factor estimator will be minimax. A specific localized shrinkage factor estimator is described. The nonapplicability of the derived condition to some estimators is (unfortunately) shown. In the last section several comments concerning these results are made.

NASA Project: "An Empirical Bayes Approach to Some Spatial Analysis Problems with Special Attention to Remotely Sensed Satellite Imagery".

1. Minimax Results Applicable to Localized Shrinkage Factor Estimators.

We shall be considering the equal, known variance case

$$Y|\theta \sim N_k(\theta, I), \qquad (1)$$

for assessing estimators $\hat{\theta}$ of θ with respect to SEL,

$$L(\theta,\hat{\theta}) = \Sigma(\hat{\theta}_i-\theta_i)^2. \qquad (2)$$

Stein (1981 Annals) shows that for fairly general estimators of the form

$$\hat{\theta} = Y + g(Y), \qquad (3)$$

$g:R^k \to R^k$ (see Stein (1981) for exact conditions on g), the following results hold:

$$E(Y_i + g_i(Y) - \theta_i)^2 = 1 + E(g_i^2(Y) + 2\nabla_i\, g_i(Y)) \qquad (4)$$

and so

$$E\| Y + g(Y) - \theta\|^2 = k + E(\|g(Y)\|^2 + 2\nabla.g(Y)) \qquad (5)$$

Here

$$\nabla_i = \frac{\partial}{\partial Y_i} \qquad (6)$$

and

$$\nabla.g(Y) = \Sigma\nabla_i g_i(Y) \qquad (7)$$

where $g_i(Y)$ is the i^{th} component of g.

This result may be applied to estimators of the form

$$\hat{\theta} = Y - A\,[\lambda(Y)]\,AY \qquad (8)$$

where A is a preassigned matrix, $\lambda: R^k \to R^k$, and $\Lambda[\lambda(Y)]$ is the kxk diagonal matrix with diagonal elements $\lambda_1(Y), \ldots, \lambda_k(Y)$. $\lambda_i(Y)$ is the localized shrinkage factor for Y_i. Thus

$$\hat{\theta}_i = Y_i - \lambda_i(Y)A_iY \tag{9}$$

where A_i is the i^{th} row of A (the i^{th} column of A will be denoted a_i and a_{ij} will represent the ij^{th} element of A). We shall assume that $\lambda(Y)$ is chosen so that the necessary expectations exist

Lemma 1.

$$E(Y_i - \lambda_i(Y)A_iY - \theta_i)^2 =$$
$$1 + E(\lambda_i^2(Y)\ (A_iY)^2 - 2\lambda_i(Y)\ a_{ii} - 2A_iY\nabla_i\lambda_i(Y)) \tag{10}$$

proof: Apply Stein's result with

$$g_i(Y) = -\lambda_i(Y)A_iY,$$

so
$$g_i^2(Y) = \lambda_i^2(Y)\ (A_iY)^2$$

and
$$\nabla_i g_i(Y) = -\lambda\ (Y)\ a_{ii} - (A_iY)(\nabla_i\lambda_i(Y)).\quad \text{QED.}$$

If $\lambda_i(Y)$ is of the form

$$\lambda_i(Y) = \frac{1}{Y'B^1Y} \tag{11}$$

where B^1 is a positive definite (symmetric) matrix, then

$$\nabla_i\lambda_i(Y) = -2\lambda_i^2(Y)\ Y'b_i^1. \tag{12}$$

Theorem 1. For $\lambda_i(Y)$ as in (11),

$$E(Y_i - \lambda_i(Y)A_iY - \theta_i)^2 =$$

$$1 + E(\lambda_i^2(Y)\ Y'(A_i'A_i - 2a_{ii}B^i + 4\ b_i^iA_i)Y). \qquad (13)$$

Hence $\hat{\theta}$ is minimax if

$$C^i = 2a_{ii}B^i - A_i'A_i - 4b_i^iA_i \geq 0 \qquad (14)$$

(is n.n.d.) for i=1, ..., k.

proof: Applying Lemma 1 when (12) holds yields

$$E(Y_i - \lambda_i(Y)A_iY - \theta_i)^2 = 1 - E(\lambda_i^2(Y)\ Y'C^iY)$$

and so if $C^i \geq 0$ we have

$$E\|Y - A[\lambda(Y)]AY - \theta\|^2 \leq k. \quad \text{QED.}$$

2. A Proposed Localized Shrinkage Factor Estimator

Rewriting (9) as

$$\hat{\theta}_i = (1 - \lambda_i(Y))Y_i + \lambda_i(Y)\ A_i^*Y \qquad (15)$$

where $A^* = I-A$, it can be seen that $\lambda_i(Y)$ determines the degree of shrinkage from Y_i to A_i^*Y. Let

$$\lambda_i(Y) = \frac{d_i}{\sum_j (a_{ij}^*)^2 (Y_j - A_i^*Y)^2}, \qquad (16)$$

where d_i is a positive constant. This choice of $\lambda_i(Y)$ allows the shrinkage at Y_i to be determined by the Y_j's which have nonzero weight in A_i^*Y.

Lemma 2. $$\sum_j (a^*_{ij})^2 (Y_j - A^*_i Y)^2 = d_i Y' B^i Y \quad \text{for}$$

$$d_i B^i = \sum_j [a^*_{ij}(I_j - A^*_i)]' [a^*_{ij}(I_j - A^*_i)] \tag{17}$$

where I_j is the j^{th} row of the identity matrix I.
In terms of A this is

$$\begin{aligned} d_i B^i = {} & [(1-a_{ii})^2 + \sum_{j\neq i} a_{ij}^2] A_i' A_i \\ & + \sum_{j\neq i} a_{ij}^2 (A_i'(I_j - I_i) + (I_j - I_i)' A_i \\ & + (I_j - I_i)'(I_j - I_i)). \end{aligned} \tag{18}$$

proof: (17) follows upon writing

$$\sum_j (a^*_{ij})^2 (Y_j - A^*_i Y)^2 = \sum_j [a^*_{ij}(I_j - A^*_i) Y]^2 .$$

Noting

$$\begin{aligned} a^*_{ij}(I_j - A^*_i) &= (1-a_{ii}) A_i && \text{for } i=j \\ &= -a_{ij}(A_i + (I_j - I_i)) && \text{for } i\neq j \end{aligned}$$

and expanding (17) yields (18). QED

The estimator $\hat{\theta}$ with $\lambda_i(Y)$ defined in (16) with

$$a^*_{ij} = \bar{a}^*_i . \tag{19}$$

for j such that a^*_{ij} is nonzero, where $\bar{a}^*_{ij}$ is the average of the nonzero a^*_{ij}, is a spatially moving version of the James-Stein estimator.

Recall that in order to apply Theorem 1 B^1 must be positive definite (required for the expectations in (13) to exist). From (17) it can be seen that if $a^s_{ik} = 0$ then the k^{th} row (and column) of B^1 will consist entirely of zeros, and so B^1 would not be positive definite.

3. Comments.

For spatial data which exhibits only local continuity, localized shrinkage factor estimators can reasonably be expected to do better with respect to MSE than estimators involving only a single (global) shrinkage factor. A number of simulations have shown this to be the case.

Theorem 1 provides sufficient conditions for showing that estimators with localized shrinkage factors are minimax but, as shown in Section 2, these are probably too restrictive. Can the requirement that $B>0$ be eased? to $B\geq 0$? Perhaps a different approach is necessary.

OMIT TO P125

Covariance Hypotheses For LANDSAT Data

Charles Peters
Department of Mathematics
University of Houston

COVARIANCE HYPOTHESES FOR LANDSAT DATA

1. Introduction

Population considered — All "fields" in a 2-dimensional digital image containing a particular number N of pixels and from a particular objective class.

$\mathbb{X} = (\mathbb{X}_1 | \cdots | \mathbb{X}_N)_{d \times N}$ — data from a sampled field. The columns are the measurements from individual pixels arranged in dictionary ordering by line and column no. in the image.

$\mathbb{X}_i$ and $\mathbb{X}_{i+1}$ probably come from geographically adjacent areas, so they are not independent. Assume that $\{\mathbb{X}_i\}_{i=1}^{N}$ is stationary with covariance function $\Gamma(h) = \operatorname{cov}(\mathbb{X}_i, \mathbb{X}_{i+h})$ and that $\mathbb{X}$ is normally distributed in $\mathbb{R}^{Nd}$.

2. Covariance Hypotheses

H_1 : $\Gamma(h) = \Omega^{\frac{1}{2}} A^{|h|} \Omega^{1/2}$ where

$\Omega_{d \times d}$ is positive definite

$A_{d \times d}$ is symmetric with $\rho(A) < 1$.

Equivalently,

$$X_{i+1} - \mu = B(X_i - \mu) + \epsilon_i$$

$$B = \Omega^{1/2} A \, \Omega^{-1/2}$$

$\epsilon_1, \cdots, \epsilon_{N-1}$ independently dist. as $N_d(0, (I - B^2)\Omega)$.

$\mu = \mathcal{E}[X_i]$ is unknown

H_2 : $\Gamma(h) = \begin{cases} \Sigma & \text{if } h \neq 0 \\ \psi + \Sigma & \text{if } h = 0, \end{cases}$

where ψ and $\psi + N\Sigma$ are positive definite.

Equivalently, $X_1, \cdots, X_N$ are exchangeable;

$X Q \underset{dist.}{=} X$ for each $N \times N$ permutation matrix

Theorem 1: H_2 is the most general covariance hypothesis under which $(\bar{X}, S)$ is a sufficient statistic, where $\bar{X}$ and S are the mean and scatter of $X_1, \cdots, X_N$. If H_2 is true then

(a) $XW \underset{dist.}{=} \bar{X}$ for each W in the group

$$\Theta_N' = \{W \mid W_{N \times N} \text{ is orthogonal and } WJ_N = J_N\}$$

where $J_N^T = (1, 1, \cdots, 1)$;

(b) if $P_{N \times (N-1)}$ satisfies $P^T P = I_{N-1}$ and $P^T J_N = 0$ then $Y = XP$ has columns $Y_1, \cdots, Y_{N-1}$ which are identifically distributed as $N_d(0, \psi)$; $\bar{X}$ and S are independent: $\bar{X} \sim N_d(\mu, \Sigma + \frac{1}{N}\psi)$, $S \sim W_d(N-1, \psi)$.

Theorem 2: Assuming only (a) of Theorem 1 (without normality) the distribution of

$$\frac{N-d-2}{d} Y_1^T \left(\sum_{j=2}^{N-1} Y_j Y_j^T \right)^{-1} Y_1$$

is central $F_{d, N-d-2}$ where $Y = XP$ as in (b). (A. P. Dawid, 1977)

3. Approximating H_1 by H_2

Let $f(x)$ be a ~~dense~~ normal density on $\mathbb{R}^{dN}$ with parameters μ, Ω, A satisfying H_1.

We want a density $\hat{f}(x)$ satisfying H_2 ~~wh~~ with parameters ν (the mean), Ψ and Σ which approximates, in a sense, $f(x)$.

Criterion - choose $\hat{f}$ to minimize the relative entropy

$$H(\hat{f},f) = \int_{\mathbb{R}^{Nd}} f \log\left(\frac{f}{\hat{f}}\right) \quad .$$

The sharpest relationship we can find between H and the $\mathcal{L}_1$ distance is the following.

Theorem 3: For any densities $f, \hat{f}$ and any $\epsilon > 0$

$$\frac{1}{2}\int_{\mathbb{R}^{Nd}} |f-\hat{f}| < \epsilon + g(\epsilon)\, H(\hat{f},f) \quad \text{where}$$

$$g(\epsilon) = \frac{\epsilon}{\epsilon - \log(1+\epsilon)} .$$

Theorem 4: If f satisfies H_1, the $\hat{f}$ satisfying H_2 which minimizes $H(\hat{f},f)$ is the one for which

$$\nu = \mu$$

$$\psi = \frac{N}{N-1}\Omega - \frac{1}{N-1}\Omega^{\frac{1}{2}}\left[(I-A)^{-1}(I+A) - \frac{2}{N}A(I-A)^{-2}(I-A^N)\right]\Omega^{\frac{1}{2}}$$

$$\Sigma = \frac{1}{N-1}\Omega^{\frac{1}{2}}\left[(I-A)^{-1}(I+A) - \frac{2}{N}A(I-A)^{-2}(I-A^N)\right]\Omega^{\frac{1}{2}} - \frac{1}{N-1}\Omega .$$

Note: $R = \psi + N\Sigma$ is nearly constant for large N.

For large N, $\max H(\hat{f},f) \approx -\frac{N-1}{2}\log|I-A^2|$.

4. Mixture Density Estimation

Let k fields be sampled from a population representing m classes in proportions $\alpha_1,\cdots,\alpha_m$.

Field sizes are $N_1,\cdots,N_k$, not all the same, but restricted so that N_i is independent of the data (except to specify how much.) and also of the field's class identity.

The data matrices are $X^1, \cdots, X^k$ (X^i is $d \times N_i$).

$f(x \mid \ell, N)$ is the density of X given that the field comes from class ℓ and has size N.

$f(x \mid \ell, N)$ is of the parametric form of H_1 or H_2.

Problem - Estimate the α_ℓ's and the densities $f(x \mid \ell, N)$.
with a sample of unlabelled fields.

H_1 is probably more realistic than H_2, but is much harder to estimate in a mixture setting.

H_2 is also very difficult unless for each class the parameters ψ_ℓ, Σ_ℓ satisfy $\psi_\ell + N\Sigma_\ell$ = constant, independent of N. Letting $\Omega_\ell = \psi_\ell + N\Sigma_\ell$ the parameters $\alpha_\ell, \mu_\ell, \psi_\ell, \Omega_\ell$ are relatively easy to estimate.

The assumption $\psi_\ell + N\Sigma_\ell = \Omega_\ell$ is justified (for large N) only because we want H_2 to approximate H_1.

5. Testing H_2

Table 1 shows the distribution of the F-ratios described in Theorem 2 for 216 fields from LACIE segment 1645 and 57 fields from segment 1633. The χ^2's given are significant at between 10% and 20%.

The "fields" are those found by AMOEBA.

TABLE 1 - Distribution of F-Ratios

Segment 1645 - 216 Fields

Percentiles	0 - 5%	5 - 10%	10 - 90%	90 - 95%	95 - 100%
Number	18	14	163	9	12
Frequency	8.2%	6.5%	75.5%	4.2%	5.6%

$X^2 - 6.72$

Segment 1633 - 57 Fields

Percentiles	0 - 5%	5 - 10%	10 - 90%	90 - 95%	95 - 100%
Number	6	1	44	4	2
Frequency	10.5%	1.3%	77.7%	7.0%	3.5%

$X^2 = 5.45$

N83
23071

UNCLAS

This Page Intentionally Left Blank

N83 23071 D6

COVARIANCE HYPOTHESES FOR LANDSAT DATA

by

Henry P. Decell
and
Charles Peters
Department of Mathematics
University of Houston

ABSTRACT

Two covariance hypotheses are considered for LANDSAT data acquired by sampling "fields", one an autoregressive covariance structure and the other the hypothesis of exchangeability. A minimum entropy approximation of the first structure by the second is derived and shown to have desireable properties for incorporation into a mixture density estimation procedure. Results of a rough test of the exchangeability hypothesis are presented.

I. Introduction.

Let $X = (X_1/\cdots/X_n)$ be a random $d \times N$ matrix having a normal distribution in R^{dN}. In the application we have in mind the columns X_i of X are the multispectral measurements from the set of pixels in a field randomly chosen from the set of all fields of a particular size N and in a particular crop class. The indexing of the X_i designates the dictionary ordering by line and column number in the image. Since there is a high probability that X_i and X_{i+1} come from spatially adjacent pixels, the columns of X are not epected to be independent. We will consider two hypotheses concerning the covariance of the X_i's. In each, the process $\{X_i\}_{i=1}^{N}$ is stationary with unknown mean μ and covariance function $\Gamma(h) = \text{cov}\,(X_i, X_{i+h})$.

H1: $\{X_i\}$ is first order autoregressive with $\Gamma(h) = \Omega^{\frac{1}{2}}A^{|h|}\Omega^{\frac{1}{2}}$, where Ω_{dxd} is positive definite and A_{dxd} is symmetric with spectral radius less than 1. That is,

$$X_{i+1} - \mu = B(X_i - \mu) + \epsilon_i$$

where $B = \Omega^{\frac{1}{2}}A\Omega^{-\frac{1}{2}}$ and $\epsilon_1, \cdots, \epsilon_{N-1}$ are independently normally distributed with mean 0 and variance-covariance matrix $(I - B^2)\Omega$.

H2: The r.v's $X_1, \cdots, X_N$ are exchangeable; i.e., the distribution of XQ is the same as that of X for each $N \times N$ permutation matrix Q. In this case, $\Gamma(0) = \psi + \Sigma$ and $\Gamma(h) = \Sigma$ for $h \neq 0$, where ψ and $\psi + N\Sigma$ are $d \times d$ positive definite symmetric matrices.

H2 implies a number of things about the distribution of X, some of which are listed below.

Theorem 1. If the distribution of X is normal and satisfies H2, then

(a) the distribution of XW is the same or that of X for each W in the group $O_N' = \{W \mid W_{N\times N}$ is orthogonal and $WJ_N = J_N\}$, where $J_N^T = (1,\cdots,1)_{1\times N}$;

(b) if $P_{N\times(N-1)}$ satisfies $P^TP = I_{N-1}$ and $P^TJ_N = 0$, then $y = XP$ has columns $y_1,\cdots,y_{N-1}$ which are independently distributed as $N_d(0,\psi)$; furthermore the statistics $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ and $S = \sum_{i=1}^{N}(X_i - \overline{X})(X_i - \overline{X})^T$ are independently distributed as $N_d(\mu,\Sigma + \frac{1}{N}\psi)$ and $W_d(N-1,\psi)$ respectively; (If $\mathcal{N}$ is a family of normal distribution of X containing $N_{dN}(0,I)$, then $(\overline{X},S)$ is sufficient for $\mathcal{N}$ if and only if each member of $\mathcal{N}$ satisfies H2).

Proof: Both (a) and (b) are easily obtained after writing the covariance of X as $\Gamma = \psi \otimes I_N + \Sigma \otimes J_NJ_N^T$. The ~~first~~ last assertion is proved in [ ], and depends on (a) and the fact that $(\overline{X},S)$ is a maximal invariant of O_N' acting on X as it does.

2. Approximating H1 by H2.

Suppose the density $f(x)$ of X actually satisfies H1 with parameter values μ, Ω and A. In this section we will show that , in a sense, the lust [best] approximation to the distribution of X by one satisfying H2 is obtained when Σ is nearly proportional to $1/N$, for large N. This is plausible, since the average covariance between pairs of distinct columns under the Markov assumption is $O(\frac{1}{N})$, see [ ]. The method of approximation we choose is to minimize the relative entropy

$$H(\hat{f}, f) = \int_{R^{dN}} \left(\log \frac{f(x)}{\hat{f}(x)}\right) f(x)dx$$

where $\hat{f}(x)$ is a density function satisfying H_2 with parameters ν (the mean), ψ and Σ. The entropy $H(\hat{f}, f)$ is weakly related to the L_1 distance by the following inequality, which is the sharpest we have been able to find. The proof is essentially that given by Geman [ ] in proving that $H(f_n, f) \to 0$ implies that $\int_{-\infty}^{\infty} |f - f_n| \to 0$.

<u>Theorem 2</u>: Let $\hat{f}$ and f be arbitrary density functions on R^k and let $\epsilon > 0$. Then

$$\frac{1}{2}\int |\hat{f} - f|\, dx \le \epsilon + \frac{\epsilon}{\epsilon - \log(1+\epsilon)} H(\hat{f}, f).$$

<u>Proof:</u> Define $g(\epsilon)$ for $\epsilon > 0$ by $g(\epsilon) = \frac{\epsilon}{\epsilon - \log(1+\epsilon)}$. Then g is positive and strictly decreasing on $(0,\infty)$. Therefore, for $\frac{\hat{f}}{f} - 1 > \epsilon$,

$$\frac{\hat{f}}{f} - 1 < g(\epsilon)\left[\frac{\hat{f}}{f} - 1 - \log\frac{\hat{f}}{f}\right] \quad \text{and}$$

$$\frac{1}{2}\int_{R^n} |\hat{f} - f| = \int_{\hat{f} > f} (\hat{f} - f) = \int_{0 < \frac{\hat{f}}{f} - 1 \le \epsilon} \left(\frac{\hat{f}}{f} - 1\right)f + \int_{\frac{\hat{f}}{f} - 1 > \epsilon} \left(\frac{\hat{f}}{f} - 1\right)f$$

$$\le \int_{0 < \frac{\hat{f}}{f} - 1 \le \epsilon} \left(\frac{\hat{f}}{f} - 1\right)f + g(\epsilon) \int_{\frac{\hat{f}}{f} - 1 > \epsilon} \left[\frac{\hat{f}}{f} - 1 - \log\frac{\hat{f}}{f}\right] f$$

$$\le \epsilon + g(\epsilon) \int_{R^k} \left[\frac{\hat{f}}{f} - 1 - \log\frac{\hat{f}}{f}\right] f$$

$$= \epsilon - g(\epsilon) \int_{R^k} (\log \frac{\hat{f}}{f}) f$$

$$= \epsilon + g(\epsilon) H(\hat{f}, f). \qquad \text{QED.}$$

Lemma 1: If $X = (X_1 | \ldots | X_N)_{d \times N}$ satisfies H1, and $\overline{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$,

$$S = \sum_{j=1}^{N} (X_i - \overline{X})(\overline{X}_i - \overline{X})^T$$

then (a) $\mathcal{E}(\overline{X}) = \mu$

(b) $cov(\overline{X}) = \frac{1}{N} \Omega^{\frac{1}{2}} [(I - A)^{-1}(I + A)^{-2} A(I - A^N)] \Omega^{\frac{1}{2}}$

(c) $\mathcal{E}(S) = N\Omega - \Omega^{\frac{1}{2}} [(I - A)^{-1}(I + A) - \frac{2}{N}(I - A)^{-2} A(I - A^N)] \Omega^{\frac{1}{2}}$

Under H2, the log likelihood function is

$$\log \hat{f}^{\Gamma}(x) = -\frac{N-1}{2} \log|\psi| - \frac{1}{2} \log|R| - \frac{1}{2} \operatorname{tr} \psi^{-1} S$$

$$- \frac{N}{2} \operatorname{tr} R^{-1} (\overline{X} - \nu)(\overline{X} - \nu)^T ,$$

when $R = \psi + N\Sigma$. By taking expectations and then differentials with respect to the parameters, one sees that the maximum of $\mathcal{E}_f(\log \hat{f})$ is achieved when

$$\nu = \mathcal{E}_f(\overline{X})$$

$$\psi + \Sigma = cov_f(\overline{X}) + \frac{1}{N} \mathcal{E}_f(S)$$

$$\Sigma = cov_f(\overline{X}) - \frac{1}{N(N-1)} \mathcal{E}_f(S)$$

Combining these results with Lemma 1, one has the following theorem.

Theorem 3. If f in a dN-variate normal density satisfying H1, then the normal density $\hat{f}$ satisfying H2 which minimizes $H(\hat{f}, f)$ is the ~~our~~ one for which

$$\nu = \mu$$

$$\psi = \frac{N}{N-1}\Omega - \frac{1}{N-1}\Omega^{\frac{1}{2}}\left[(I - A)^{-1}(I + A) - \frac{2}{N}A(I - A)^{-2}(I - A^N)\right]\Omega^{\frac{1}{2}}$$

$$\Sigma = \frac{1}{N-1}\Omega^{\frac{1}{2}}\left[(I - A)^{-1}(I + A) - \frac{2}{N}A(I - A)^{-2}(I - A^N)\right]\Omega^{\frac{1}{2}} - \frac{1}{N-1}\Omega$$

Notice that $\psi + \Sigma = \Omega$ and that $R = \psi + N\Sigma$ $= \Omega^{\frac{1}{2}}\left[(I - A)^{-1}(I + A) - \frac{2}{N}A(I - A)^{-2}(I - A^N)\right]\Omega^{\frac{1}{2}}$ is always positive definite and is effectively independent of N for large N. The corresponding maximum value of $\mathcal{E}_f(\log \hat{f})$ is

$$\mathcal{E}_f(\log \hat{f}) = -\frac{N-1}{2}\log|\psi| - \frac{1}{2}\log|R| - \frac{Nd}{2}.$$

For large values of N, this is nearly

$$\mathcal{E}_f(\log \hat{f}) \approx -\frac{N}{2}\log|\Omega| - \frac{1}{2}\log|(I - A)^{-1}(I + A)| - \frac{Nd}{2}.$$

Under H1,

$$\log f(X) = -\frac{N}{2}\log|\Omega| - \frac{N-1}{2}\log|I - A^2| - \frac{1}{2}Q(X),$$

where

$$Q(X) = \operatorname{tr}\Omega^{-\frac{1}{2}}(I - A)^{-1}\Omega^{-\frac{1}{2}}(X_1 - \mu)(X_1 - \mu)^T + \operatorname{tr}\Omega^{-\frac{1}{2}}(I - A^2)^{-1}\Omega^{-\frac{1}{2}}(X_N - \mu)(X - \mu)^T$$

$$- 2\sum_{j=1}^{N-1}\operatorname{tr}\Omega^{-\frac{1}{2}}A(I - A^2)^{-1}\Omega^{-\frac{1}{2}}(X_{j+1} - \mu)(X_j - \mu)^T$$

$$+ \sum_{j=2}^{N-1} \mathrm{tr}\, \Omega^{-\frac{1}{2}} (I - A^2)^{-1} (I + A^2)\, \Omega^{-\frac{1}{2}} (X_i - \mu)(X_i - \mu)^T .$$

Hence,

$$\mathcal{E}_f(\log f) = - \frac{N}{2} \log|\Omega| - \frac{N-1}{2} \log|I - A^2| - \mathrm{tr}(I - A^2)^{-1} A^2$$

$$+ (N-1)\mathrm{tr}(I - A^2)^{-1} A \quad - \frac{N-2}{2} \mathrm{tr}\,(I - A^2)^{-1}(I + A^2)$$

$$= - \frac{N}{2} \log|\Omega| - \frac{N-1}{2} \log|I - A^2| - \frac{Nd}{2} .$$

Therefore, for large N, the minimum relative entropy is

$$H(\hat{f}, f) \approx - \frac{N-1}{2} \log|I - A^2| \;+ \frac{1}{2} \log |(I - A)^{-1}(I + A)| .$$

One might think that because the inequality in Theorem 2 is symmetric with respect to f and $\hat{f}$, the $\hat{f}$ minimizing $H(f, \hat{f})$ should also be investigated. Unfortunately, $H(f, \hat{f})$ does not seem to have a minimum for all values of Ω, A, μ. We do not know if $H(f, \hat{f})$ can be made smaller than the minimum value of $H(\hat{f}, f)$.

3. A Mixture Density Model for LANDSAT Data.

Suppose K fields are sampled from a population of fields representing m crop classes in proportion $\alpha_1, \ldots, \alpha_m$ (these are not areal proportions; rather, they are the probabilities that a randomly ~~related~~ selected field from the population will belong to the given classes.) The sizes $N_1, \ldots, N_k$ of the sampled fields will vary; however, we may suppose that the population is suitably restricted so that each N_j is independent of the crop class and spectral data from the associated field, except to determine the dimensions of the data matrix corresponding to that field. Let $n_1, \ldots, n_k$

and $x^1, \cdots, x^k$ be the observed field sizes and data matrices and suppose the classification of each field is unknown. Then the log-likelihood function for the given observations is

$$l = \sum_{j=1}^{K} \log P[N=n_j] + \sum_{j=1}^{K} \log \sum_{l=1}^{m} \alpha_l f_l(x^j \mid N=n_j)$$

where $f_i(x^j \mid N=n_j)$ is the density of X^j given that field$_j$ is in the $l^{\underline{th}}$ class. The density f_i is of the parametric form associated either with H1 or H2. We are interested in estimating the parameters (particularly the α_i's) by maximum likelihood. Although he have almost no empirical basis for believing so, we consider H1 more realistic than H2; however, we know of no computationally efficient way of maximizing the likelihood function l under H1 (see Fuller [ ] for a discussion of the difficultied involved for even a single class). Even with H2, the likelihood equations are difficult to solve and the EM algorithm has no simple formulation, unless one assumes that for each class the parameter Σ is a constant times $\frac{1}{N}$. In this case, the likelihood equation take on a simple form and are easy to solve iteratively [ ]. If H2 is taken seriously, as in the random effects model suggested by Feiveson (see [ ]) there is no justification for the additional assumption that $\Sigma = \frac{\Sigma_0}{N}$. If, however, one regards H1 as realistic and H2 as purely the most general feasible covariance model for purposes of estimating the α_i's, then the discussion in section 2 shows that the additional assumption $\Sigma = \frac{\Sigma_0}{N}$ is reasonable for approximating the true densities, at least for large field sizes $N_1, \cdots, N_k$.

4. Testing the Hypotheses.

Hypotheses and H2 will be subjected to several tests and the results discussed in a future report. We remark that likelihood ratio test for H1 and H2 against all normal alternatives would be difficult to implement because of sampling difficulties, large dimensionality, and the aforementioned problems in obtaining MLEs under H1. We have conducted an informal test of H2 based on part(b) of Theorem 1. Let the data matices from a sample of k fields be $X^1, \cdots, X^k$ having dimensions $d \times N_1, \cdots, d \times N_k$. Let $P_1, \cdots, P_k$ be any $N_i \times (N_i - 1)$ matrices satisfying the conditions of Theorem 1(b). Let $Y^i = X^i P_i = (y_1^i | \cdots | y_{N_i-1}^i)$ and let $F_i = \frac{N_i - d - 1}{d} (y_1^i)^T [\sum_{j=2}^{N_i-1} (y_j^i)(y_j^i)^T]^{-1} y_1^i$. Then F_i has a central F distribution with d and $N_i - d - 1$ degrees of freedom. Under the hypothesis H3 stated in part (a) of Theorem 1, this result is entirely independent, of the distribution of X^i, and follows from results of David [ ] showing that the exact distribution of F_i depends only on the right spherical symmetry of Y^i. However, H3 is stronger, in general than H2, except under normality. Table 1 shows the number of F_is which fell in the upper and lower tails of their respective F distributions and the associated X^2 statistics for 216 fields from LACIE segment 1645 and 57 fields from segment 1633. In point of fact, the "fields" represented in Table 1 are those produced by an automatic image segmentation program (AMOEBA) and may not be representative of real agricultural fields. The X^2 statistics are significant at levels between 10% and 20% so that Table 1 provides rather weak disconfirmation of H2.

This Page Intentionally Left Blank

OMIT TO P.123

A Hypothesis Test for the Rank of the Minimal Linear Sufficient Statistic

Richard A. Redner
and
William A. Coberly

Department of Mathematical Science
University of Tulsa

January 28, 1983

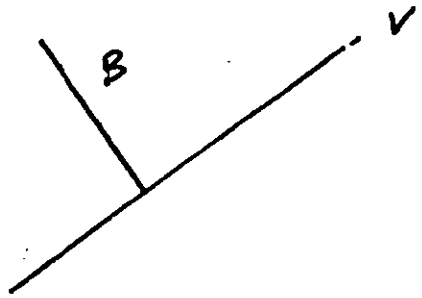
DEFINITION LET $\{f_\theta\}$ $\theta \in \Theta$ BE A FAMILY OF DENSITY FUNCTIONS ON R^n WITH COMMON SUPPORT. A MAPPING $T: R^n \to R^k$ IS A SUFFICIENT STATISTIC FOR THIS FAMILY IFF FOR SOME FIXED $\theta_0 \in \Theta$

$$\frac{f_\theta(x)}{f_{\theta_0}(x)} = g \circ T(x) .$$

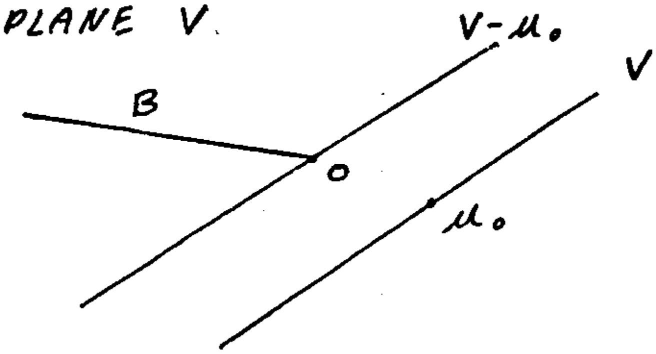
LET $y_i \sim N(\mu_i, \Omega_i)$ $i = 0, 1, \cdots, m$.

SUPPOSE THAT THE $k \times n$ MATRIX B IS A SUFFICIENT STATISTIC

1. IF $\mu_0 = 0$ AND $\Omega_i = I$ $i = 0, \cdots, m$ THEN THE MEANS LIE IN A K-DIMENSIONAL SUBSPACE V.



2. IF $\Omega_i = \Omega$ $i = 0, 1, \cdots, m$ THEN THE MEANS LIE IN K-DIMENSIONAL HYPERPLANE V.



THEOREM (PETERS, DECELL, + REDNER) A FULL RANK $K \times n$ MATRIX B IS SUFFICIENT FOR $\xi_i \sim N(\mu_i, \Omega_i)$ $i = 0, 1, \cdots, m$ IFF

$$P(\Omega_i - \Omega_0) = \Omega_i - \Omega_0$$
$$P(\mu_i - \mu_0) = \mu_i - \mu_0$$

WHERE $P = \Omega_0 B^T (B \Omega_0 B^T)^{-1} B$

NOTE: P IS A PROJECTION IN THE INNER PRODUCT

$$(X, Y) = X^T \Omega_0^{-1} Y$$

CONSIDER THE CHANGE OF VARIABLES

$$z = \Omega_0^{-1/2}(X - \mu_0)$$

THEN WE GET

$$\hat{f}_i(z) \sim N(\mu_i - \mu_0, \Omega_0^{-1/2}\Omega_i\Omega_0^{-1/2}) \quad i = 1, \cdots, m$$

$$\hat{f}_0(z) \sim N(0, I)$$

<u>COROLLARY</u> A FULL RANK $k \times n$ MATRIX B IS SUFFICIENT FOR $\{\hat{f}_i\}_{i=0}^{m}$ IFF

$$\hat{P}(\hat{\Omega}_i - I) = \hat{\Omega}_i - I \qquad i = 0, 1, \cdots, m$$
$$\hat{P}(\hat{\mu}_i) = \hat{\mu}_i$$

WHERE $\hat{P} = B^T(BB^T)^{-1}B$

COROLLARY A FULL RANK $K \times n$ MATRIX B IS SUFFICIENT FOR $\{\hat{\xi}_i\}_{i=0}^{m}$ IFF THE SPAN OF THE ROWS OF B CONTAINS $\hat{\mu}_i$ AND THE COLUMNS OF $\hat{\Omega}_i - I$ FOR $i = 1, \cdots, m$.

PROOF RANGE $(P = B^T (BB^T)^{-1} B)$

$$= \text{SPAN}\{\text{ROWS OF } B\}$$

THEOREM. A FULL RANK $K \times n$ MATRIX B IS SUFFICIENT FOR $\{\hat{\xi}_i\}_{i=0}^{m}$ IFF THERE IS A CHANGE OF VARIABLES $Y = HZ$ SO THAT

$$\tilde{\xi}_i(Y) \sim N\left(\begin{pmatrix} 0 \\ \tilde{\eta}_i \end{pmatrix}, \begin{pmatrix} I & 0 \\ 0 & \tilde{R}_i \end{pmatrix}\right)$$

WHERE H IS PURE ROTATION
$\tilde{\eta}_i$ IS $K \times 1$
$\tilde{R}_i$ IS $K \times K$ POSITIVE DEFINITE

PROOF LET H BE A ROTATION OF THE FORM

$$H = \begin{pmatrix} D^{\perp} \\ D \end{pmatrix}$$

WHERE

$$D^{\perp}(\hat{\Omega}_i - I) = 0$$
$$D^{\perp}(\hat{u}_i) = 0$$

THEN $$H\hat{u}_i = \begin{pmatrix} D^{\perp}\hat{u}_i \\ D\hat{u}_i \end{pmatrix} = \begin{pmatrix} 0 \\ \tilde{n}_i \end{pmatrix}$$

$$H\hat{\Omega}_i H^T = H(\hat{\Omega}_i - I)H^T + I = \begin{pmatrix} D^{\perp} \\ D \end{pmatrix}(\hat{\Omega}_i - I)\begin{pmatrix} D^{\perp} \\ D \end{pmatrix}^T + I$$

$$= \begin{pmatrix} D^{\perp}(\hat{\Omega}_i - I)D^{\perp T} & D^{\perp}(\hat{\Omega}_i - I)D^T \\ D(\hat{\Omega}_i - I)D^{\perp T} & D(\hat{\Omega}_i - I)D^T \end{pmatrix} + I$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & R_i \end{pmatrix} + I = \begin{pmatrix} I & 0 \\ 0 & \tilde{R}_i \end{pmatrix}$$

<u>COROLLARY</u> A FULL RANK $K \times n$ MATRIX IS SUFFICIENT FOR $\{\mathfrak{f}_i\}_{i=0}^{m}$ IFF THERE IS A CHANGE OF VARIABLES

$$z = H \Gamma^{-1/2} (X - \gamma)$$

SO THAT

$$\hat{\mathfrak{f}}_i (z) = N \left(\begin{matrix} 0 \\ n_i \end{matrix} , \begin{pmatrix} I & 0 \\ 0 & R_i \end{pmatrix} \right)$$

$$i = 0, 1, \cdots, m$$

AND WHERE H IS A ROTATION

NOTE: W.L.O.G ASSUME B HAS ORTHOGONAL ROWS

TEST OF HYPOTHESIS

GIVIN $\{\{X_{ij}\}_{j=1}^{N_i}\}_{i=0}^{m}$ WE DERIVE THE LIKELIHOOD RATIO CRITERION FOR TESTING THE HYPOTESIS

H_0: $\exists$ A RANK K LINEAR SUFF. STATISTIC

THE TEST STATISTIC WILL BY

$$Y = -2 \ln \frac{\sup_{H_0} L(\{\mu_i, \Lambda_i\}_{i=0}^{m})}{\sup L(\{\mu_i, \Lambda_i\}_{i=0}^{m})}$$

WHERE Y IS ASSYMPTOTICALLY χ^2.

OBSERVE THAT IF WE HAVE A RANK K LINEAR SUFFICIENT STATISTIC THEN

$$\mu_i = \gamma + \Gamma^{1/2} H^T \begin{pmatrix} 0 \\ n_i \end{pmatrix}$$

$$\Lambda_i = \Gamma^{1/2} H^T \begin{pmatrix} I & 0 \\ 0 & R_i \end{pmatrix} H \Gamma^{1/2}$$

WHERE $R_0 = I$

$H = \begin{pmatrix} \check{B} \\ B \end{pmatrix}$ IS A ROTATION

B IS SUFFICIENT

THE LOG LIKELIHOOD FUNCTION IS

$$L = -\frac{1}{2}\sum_{i=0}^{m} N_i \ln|\Omega_i| - \frac{1}{2}\sum_{i=0}^{m} N_i \, tr(\Omega_i^{-1} S_i)$$

$$-\frac{1}{2}\sum_{i=0}^{m} N_i \, tr\, \Omega_i^{-1}(\bar{X}_i - \mu_i)(\bar{X}_i - \mu_i)^T$$

WHERE $S_i = \frac{1}{N_i}\sum_{j=1}^{N_i}(X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)^T$

$$\bar{X}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} X_{ij}$$

SUBSTITUTE $\mu_i = \gamma + \Gamma^{1/2} H^T \begin{pmatrix} 0 \\ n_i \end{pmatrix}$

$$\Omega_i = \Gamma^{1/2} H^T \begin{pmatrix} I & 0 \\ 0 & R_i \end{pmatrix} H \Gamma^{1/2}$$

OPTIMIZE OVER $\{n_i\}_{i=0}^{m}$, $R_0 = I$, $\{R_i\}_{i=1}^{m}$, γ

GIVEN $\Gamma^{1/2}$, $H = \begin{pmatrix} \tilde{B} \\ B \end{pmatrix}$

$$L(\Lambda, B, \tilde{B}) = N \ln|\Lambda| - \frac{1}{2}\sum_{i=1}^{m} N_i \ln|B\Lambda S_i \Lambda^T B^T|$$

$$- \frac{1}{2} NK - \frac{1}{2} \operatorname{tr}(\tilde{B}\Lambda \mathcal{T} \Lambda \tilde{B}^T)$$

WHERE $\mathcal{T} = WSS + BSS$

$\Lambda = \Gamma^{-1/2}$ (or $\Lambda\Lambda^T = \Gamma^{-1}$)

AND WE OPTIMIZE THIS OVER

$\Lambda, B, \tilde{B}$

NOTE: $\operatorname{tr}(\tilde{B} A \tilde{B}^T) = \operatorname{tr}(\tilde{B} A \tilde{B}^T \tilde{B} \tilde{B}^T)$

$= \operatorname{tr}(\tilde{B}^T \tilde{B} A \tilde{B}^T \tilde{B})$

$= \operatorname{tr}(I - B^T B) A (I - B^T B)$

DEGREES OF FREEDOM

1. Λ $\quad \frac{n(n+1)}{2} - 1$

2. B $\quad k(n-k)$

3 m MEANS $\{n_i\}_{i=0}^{m}$ $\quad K(m+1)$

4 GRAND MEAN γ $\quad n - K$

5 $(m-1)$ S.M $\{R_i\}_{i=1}^{m}$ $\quad (m-1)\frac{K(K+1)}{2}$

D.F $\quad$ SUM

NOTE: 2, 3, 4 GIVE

$$(K+1)n + (m-K)K$$

WHICH AGREES WITH RAO

L DEPENDS ON B UP TO THE SPAN OF THE ROWS OF B

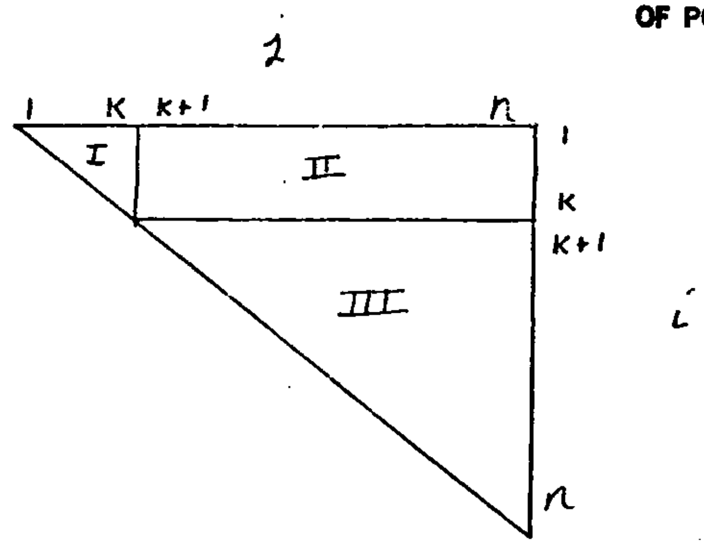
IF H IS A ROTATION

$$H = \prod_{1 \le i < j \le n} G_{ij}$$

$$G_{ij} = \begin{pmatrix} 1 & & & & & & 0 \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \cos\theta_{ij} & & & \\ & & & & \ddots & & \\ & & & & & \cos\theta_{ij} & \\ & & & & & & 1 \\ & & & & & & & \ddots \\ 0 & & & & & & & & 1 \end{pmatrix}$$

GIVENS TRANSFORMATIONS

$$B = \prod_{\substack{1 \le i \le k \\ k+1 \le j \le n}} G_{ij}$$



III $i > K$ EFFECT ONLY ROWS K+1 – n

I $j \le K$ TAKE LINEAR COMB. OF ROWS 1-K

II $1 \le i \le n$, $K+1 \le j \le n$ NEEDED TO CONSTRUCT B

STARTING VALUES

$$\Lambda \Lambda^T = \frac{1}{N_0} \sum_{j=1}^{N_0} (X_{0j} - \overline{X}_0)(X_{0j} - \overline{X}_0)^T$$

or

$$\Lambda \Lambda^T = \frac{1}{N} \sum_{i=0}^{m} N_i S_i$$

$$H = I \quad (\theta_{ij} = 0 \quad \forall ij)$$

FINALLY, IF Λ IS NOT UPDATED THEN ONLY D IS ESTIMATE.

B IS DERIVED FROM

$K(n-K)$ PARAMETERS

WHICH IS SMALL WHERE K IS SMALL OR CLOSE TO n.

This Page Intentionally Left Blank

An Adaptive Technique for Fitting LANDSAT data

by

Larry L. Schumaker and Larry F. Guseman, Jr.

Texas A&M University

Presented at NASA/MPRIA Workshop: Math/Stat

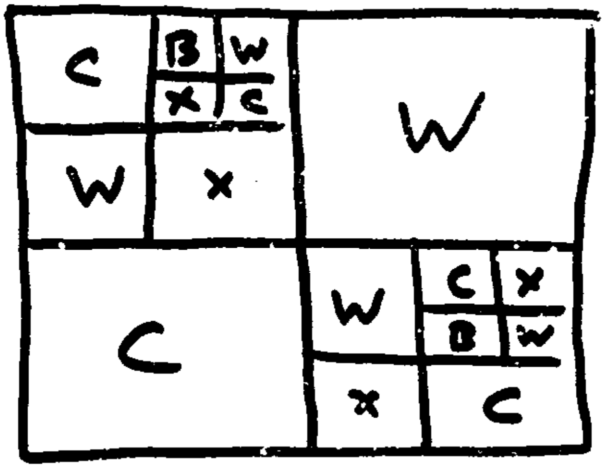
Jan. 27 - 28, 1983

Abstract

In this presentation we discuss some preliminary results on an adaptive scheme for segmenting LANDSAT images. The idea of the algorithm is to first fit a mixture of normals to the measurements in each channel to determine the number of classes represented by the data along with the mean values and variances of the measurements associated with each of these classes.

The information from the first stage is then used to adaptively compute a least-squares fit of a piecewise constant surface to the data. The resulting segmentation locates and labels the fields in the scene, and immediately yields a geometric estimation of proportions. Several numerical experiments are discussed along with a number of suggested research questions.

IDEALIZED PROBLEM



DATA ⟶ **SOLUTION**

1. number of classes
2. segmentation
3. labels
4. proportions

IDEALIZED DATA:

$$M = (M_{ij})_{1,1}^{NX\ NY}$$

where

$$M_{ij} = \begin{cases} c_1 & \text{corn} \\ c_2 & \text{wheat} \\ c_3 & \text{barley} \\ \vdots & \\ c_m & \text{bare earth} \end{cases}$$

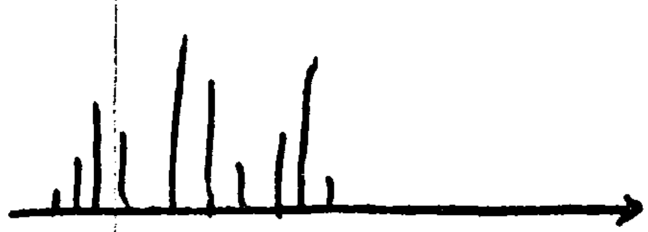
(integers) $1 \leq ! \leq 256$

REAL DATA

$$\tilde{M} = (\tilde{M}_{ij})_{1,1}^{NX\ NY}$$

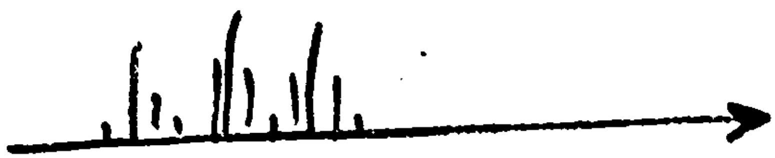
$$\tilde{M}_{ij} = M_{ij} + \theta_{ij} \quad , \ \theta_{ij} \text{ noise..}$$

HISTOGRAM



Prob. 1 Number of classes

1-D analysis of the histogram.
MIXTURE MODEL



$c_1 \; c_2 \cdots c_m$ means
$\sigma_1 \; \sigma_2 \cdots \sigma_m$ variances

rough classification
" estimate proportions

PROB 2. Segmentation.

X_1	Y_1	$size_1$	$type_1$
X_2	Y_2	$size_2$	$type_2$
⋮	⋮	⋮	⋮
X_{NF}	Y_{NF}	$size_{NF}$	$type_{NF}$

NF = number of fields.

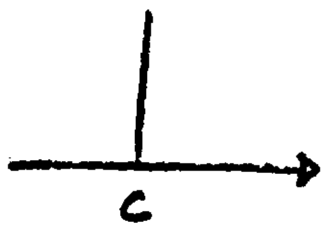
ADAPTIVE TILING:

1). Compare the mean of all DATA.

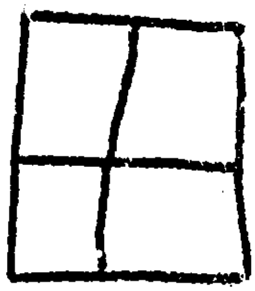
Select closest constant α from $\{C_1, \cdots, C_m\}$.

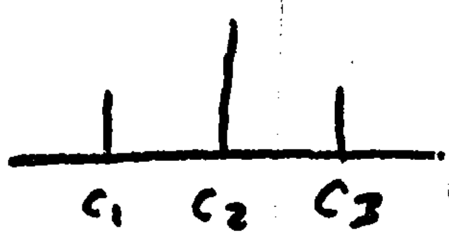
If
$$\frac{\sum\sum(\tilde{m}_{ij} - \alpha)^2}{NX \cdot NY} \leq FIT$$

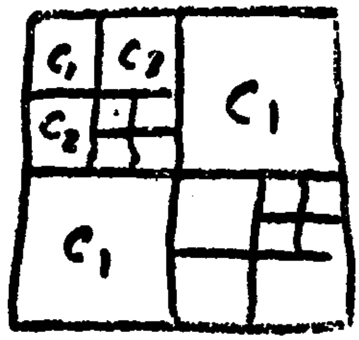
quit.



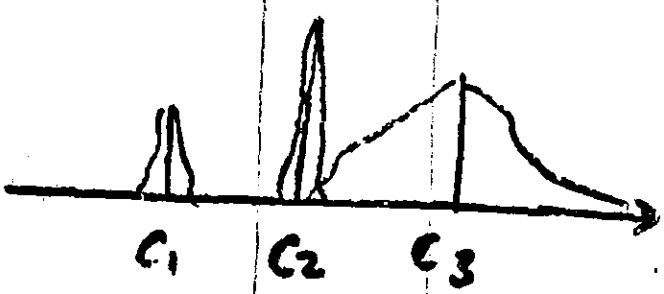
2). Split the region into 4 parts.
Fit a constant to each part.
If accuracy is Fit/4, quit, else further subdivide.







FIT controls tightness
estimate FIT from $\sigma_1, \cdot\cdot, \sigma_m$
MAY weight computation of residuals.



PROB 3. LABELS

PROB 4. PROPORTIONS

MULTICHANNEL CASE.

M_{ij} IS A VECTOR $= (M_{ij}^{v})_{v=1}^{NCH}$

M " " " $= (M^{v})_{v=1}^{NCH}$

C_i " " " $= (C_i^{v})_{v=1}^{NCH}$

σ_i " " " $= (\sigma_i^{v})$

F_{iT} " " " $= (F_{iT}^{v})$

1. Rotation of axes
2. Transformation of DATA
3. Iteration
4. Use of NON-RECTANGULAR Regions
5. Other split decision Rules
6. Parallel Processing
7. Size of SCENE.

INPUT DATA


```
11112222222222444444444444444444
11111222222222244444444444444444
11111122222222224444444444444444
11111112222222222444444444444444
11111111222222222244444444444444
11111111122222222224444444444444
11111111112222222222444444444444
11111111111222222222244444444444
11111111111122222222224444444444
11111111111112222222222444444444
11111111111111222222222244444444
11111111111111122222222224444444
11111111111111112222222222444444
11111111111111111222222222244444
11111111111111111122222222224444
11111111111111111112222222222444
```

results from ADAPT2 , initial tolerance = .40

```
11112222222222224444444444444444
11112222222222224444444444444444
11111122222222224444444444444444
11111111222222224444444444444444
11111111222222222244444444444444
11111111222222222222444444444444
11111111222222222222244444444444
11111111222222222222224444444444
11111111111122222222222244444444
11111111111122222222222244444444
11111111111111222222222244444444
11111111111111122222222444444444
11111111111111122222222224444444
11111111111111111122222222224444
11111111111111111122222222224444
11111111111111111111222222222244
```

results from ADAPT2 , initial tolerance = .10

```
11112222222222444444444444444444
11111122222222224444444444444444
11111122222222224444444444444444
11111111222222222244444444444444
11111111222222222244444444444444
11111111112222222222444444444444
11111111112222222222444444444444
11111111111122222222224444444444
11111111111122222222224444444444
11111111111111222222222244444444
11111111111111222222222244444444
11111111111111112222222222444444
11111111111111112222222222444444
11111111111111111122222222224444
11111111111111111122222222224444
11111111111111111111222222222244
```

INPUT DATA

```
11112222222222444444444444444444
11111222222222244444444444444444
11111122222222224444444444444444
11111112222222222244444444444444
11111111222222222244444444444444
11111111122222222224444444444444
11111111112222222222444444444444
11111111111222222222244444444444
11111111111122222222224444444444
11111111111112222222222444444444
11111111111111222222222244444444
11111111111111122222222224444444
11111111111111112222222222444444
11111111111111111222222222244444
11111111111111111122222222224444
11111111111111111112222222222444
```

results from ADAPT2 , initial tolerance = .80

```
11112222222222224444444444444444
11112222222222224444444444444444
11112222222222224444444444444444
11112222222222224444444444444444
11111111222222224444444444444444
11111111222222224444444444444444
11111111222222224444444444441444
11111111222222224444444444444444
11111111222222222222224444444444
11111111222222222222222244444444
11111111222222222222222244444444
11111111222222222222222244444444
11111111111111112222222244444444
11111111111111111122222222224444
11111111111111111122222222224444
11111111111111111111222222224444
```

results from ADAPT2 , initial tolerance = .60

```
11112222222222224444444444444444
11112222222222224444444444444444
11111122222222224444444444444444
11111111222222224444444444444444
11111111222222224444444444444444
11111111222222224444444444444444
11111111222222224444444444444444
11111111222222224444444444444444
11111111111122222222224444444444
11111111111122222222222244444444
11111111111111222222222244444444
11111111111111112222222244444444
11111111111111112222222224444444
11111111111111111122222222224444
11111111111111111122222222224444
11111111111111111111222222224444
```

INPUT DATA

```
3111153351111111
2311111111111111
2231111111111111
2223111111111111
2222311111111111
2222231111111444
2222223111111111
2222222311111111
2222222231111111
7722222223111111
2222222222311111
2222222222231111
2222222222223111
2222222222222311
2777122222222231
2222122222222223
```

results from ADAPT2 , initial tolerance = .90

```
3111155511111111
2311111111111111
2231111111111111
2223111111111111
2222311111111111
2222231111111111
2222223111111111
2222222311111111
2222222222222222
7722222222222222
2222222222222222
2222222222222222
2222222222222222
2222222222222222
2777222222222222
2222222222222222
```

results from ADAPT2, initial tolerance = .50

```
3111155511111111
2311111111111111
2231111111111111
2223111111111111
2222311111111111
2222231111111111
2222223111111111
2222222311111111
2222222231111111
7722222223111111
2222222222311111
2222222222231111
2222222222223111
2222222222222311
2777222222222231
2222222222222223
```

results from ADAPT2 , initial tolerance = .40

```
3111155511111111
2311111111111111
2231111111111111
2223111111111111
2222311111111111
2222231111111444
2222223111111111
2222222311111111
2222222231111111
7722222223111111
2222222222311111
2222222222231111
2222222222223111
2222222222222311
2777222222222231
```

INPUT DATA

```
12222222222222222222222222222222
11222222222222222222222222222222
11122222222222222224444222222222
11112222222222222244444422222222
11111222222222222244224422222222
11111122222222222244224422222222
12111112222222222244444422222222
12211111222222222224444222222222
11111111122222222222222222222222
11111111112222222222222222222222
11111111111222222222222211112222
11131111111122222222222211112222
11333111111112222222222222222222
11333111111111222222222222222222
11111111111111122222222222222222
11111111111111122222222222222222
```

results from ADAPT2 , initial tolerance = .70

```
11222222222222222222222222222222
11222222222222222222222222222222
11112222222222222233444422222222
11112222222222222244444422222222
11112222222222223332244332222222
11112222222222223332233322222222
11111111222222222244444422222222
11111111222222222233444422222222
11111111111111112222222222222222
11111111111111112222222222222222
11111111111111112222222222222222
11111111111111112222222222222222
11111111111111112222222222222222
11111111111111112222222222222222
11111111111111112222222222222222
11111111111111112222222222222222
```

results from ADAPT2 , initial tolerance = .30

```
11222222222222222222222222222222
11222222222222222222222222222222
11112222222222222233443322222222
11112222222222222244444422222222
11111122222222223332244222222222
11111122222222223332233222222222
11111111222222222244444422222222
11111111222222222233443322222222
11111111112222222222222222222222
11111111112222222222222222222222
11111111111122222222222222222222
11221111111122222222222222222222
11332211111111112222222222222222
11332211111111112222222222222222
11111111111111112222222222222222
11111111111111112222222222222222
```

results from ADAPT2 , initial tolerance = .20

```
11222222222222222222222222222221
11222222222222222222222222222221
11112222222222222244442222222221
11112222222222222334444332222221
11111122222222223442233442222221
11111122222222223442222442222221
11111111222222222244444422222222
11111111222222222233443322222222
11111111112222222222222222222222
11111111112222222222222222222222
11111111111122222222222222222222
11221111111122222222222222222222
11332211111111112222222222222222
11332211111111112222222222222222
11111111111111112222222222222222
```

INPUT DATA

```
33333333333333333333311111111133
13333333333333333333111111111333
11333333337333333333311111113333
11133333377733333333311133333333
11113333333733333333333133333333
11111333333333333333333333333333
11111133333333333333333333333333
21111113333333333300333333333335
22111111333333333333333333333355
12211111133333333333333333555555
22111111143333331333333313555555
21111114443333331333333355555555
11111114444433333333333555555555
11111144444444333333335555555555
11111444444444411111114444444444
11114444444444441111144444444444
```

results from ADAPT1 , initial tolerance = .75

```
33333333333333333333111111111133
22333333333333333333221111112233
11333333355333333333311111113333
11223333557733333333322113333333
11113333333333333333333333333333
11112233333333333333333333333333
11111133333333333333333333333333
11111122333333333333333333333333
11111111333333333333333333334455
11111111333333333333333333445555
11111111333333333333333333555555
11111111333333333333333344555555
11111133444433333333333555555555
11111144444433333333334455555555
11113344444444411111114444444444
11114444444444331111334444444444
```

results from ADAPT1 , initial tolerance = .70

```
33333333333333333333111111111133
33333333333333333333221111112233
33333333333333333333311111113333
33333333333333333333322113333333
33333333333333333333333333333333
33333333333333333333333333333333
33333333333333333333333333333333
33333333333333333333333333333333
11111111333333333333333333334455
11111111333333333333333333445555
11111111333333333333333333555555
11111111333333333333333344555555
11111133444433333333333555555555
11111144444433333333334455555555
11113344444444411111114444444444
11114444444444331111334444444444
```

INPUT DATA

```
33333333333333333333111111111133
13333333333333333333311111111333
11333333337333333333331111113333
11133333377733333333331113333333
11113333337333333333333313333333
11111333333333333333333333333333
11111133333333333333333333333335
21111113333333333300333333333355
22111111333333333333333333333555
12211111133333333333333333555555
22111111143333313333333333555555
21111111444333331333333335555555
11111114444433333333333555555555
11111144444444333333335555555555
11114444444444111111114444444444
11114444444444441111114444444444
```

results from ADAPT1 , initial tolerance = 1.00

```
33333333333333333333111111111133
22333333333333333333221111112233
11333333335533333333331111113333
11223333557733333333332211333333
11113333333333333333333322333333
11112233333333333333333333333333
11111133333333333333333333333344
11111122333333333300333333333355
11111111333333333333333333334455
11111111333333333333333333445555
11111111333333333333333333555555
11111111333333333333333344555555
11111133444433333333333555555555
11111144444433333333334455555555
11113344444444111111114444444444
11114444444444331111334444444444
```

results from ADAPT1 , initial tolerance = .80

```
33333333333333333333111111111133
22333333333333333333221111112233
11333333335533333333331111113333
11223333557733333333332211333333
11113333333333333333333322333333
11112233333333333333333333333333
11111133333333333333333333333344
11111122333333333300333333333355
11111111333333333333333333334455
11111111333333333333333333445555
11111111333333333333333333555555
11111111333333333333333344555555
11111133444433333333333555555555
11111144444433333333334455555555
11113344444444111111114444444444
11114444444444331111334444444444
```

INPUT DATA

```
3111155511111111
2311111111111111
2231111111111111
2223111111111111
2222311111111111
2222231111111444
2222223111111111
2222222311111111
2222222231111111
7722222223111111
2222222222311111
2222222222231111
2222222222223111
2222222222222311
2777122222222231
2222122222222223
```

results from ADAPT2 , initial tolerance = .90

```
3111155511111111
2311111111111111
2231111111111111
2223111111111111
2222311111111111
2222231111111111
2222223111111111
2222222311111111
2222222222222222
7722222222222222
2222222222222222
2222222222222222
2222222222222222
2222222222222222
2777222222222222
2222222222222222
```

results from ADAPT2, initial tolerance = .50

```
3111155511111111
2311111111111111
2231111111111111
2223111111111111
2222311111111111
2222231111111111
2222223111111111
2222222311111111
2222222231111111
7722222223111111
2222222222311111
2222222222231111
2222222222223111
2222222222222311
2777222222222231
2222222222222223
```

results from ADAPT2 , initial tolerance = .40

```
3111155511111111
2311111111111111
2231111111111111
2223111111111111
2222311111111111
2222231111111444
2222223111111111
2222222311111111
2222222231111111
7722222223111111
2222222222311111
2222222222231111
2222222222223111
2222222222222311
2777222222222231
2222222222222223
```

INPUT DATA


```
12222222222222222222222222222222
11222222222222222222222222222222
11122222222222222224444222222222
11112222222222222244444422222222
11111222222222222442224442222222
11111122222222222442222442222222
12111112222222222224444442222222
12211111122222222224444222222222
11111111122222222222222222222222
11111111112222222222222222222222
11111111111222222222222211112222
11131111111122222222222211112222
11333111111112222222222222222222
11333111111111122222222222222222
11111111111111122222222222222222
11111111111111112222222222222222
```

results from ADAPT2 , initial tolerance = .70

```
11222222222222222222222222222222
11222222222222222222222222222222
11112222222222222223344442222222
11112222222222222224444442222222
11112222222222222333224433222222
11112222222222222333223332222222
11111111222222222224444442222222
11111111222222222223344442222222
11111111111111112222222222222222
11111111111111112222222222222222
11111111111111112222222222222222
11111111111111112222222222222222
11111111111111112222222222222222
11111111111111112222222222222222
11111111111111112222222222222222
11111111111111112222222222222222
```

results from ADAPT2 , initial tolerance = .30

```
11222222222222222222222222222222
11222222222222222222222222222222
11112222222222222223443222222222
11112222222222222224444442222222
11111122222222222333322442222222
11111122222222222333322332222222
11111111222222222224414422222222
11111111222222222233443322222222
11111111112222222222222222222222
11111111112222222222222222222222
11111111111122222222222222222222
11221111111122222222222222222222
11332211111111112222222222222222
11332211111111112222222222222222
11111111111111112222222222222222
11111111111111112222222222222222
```

results from ADAPT2 , initial tolerance = .20

```
11222222222222222222222222222221
11222222222222222222222222222221
11112222222222222224444222222221
11112222222222222334444332222221
11111122222222222344223344222221
11111122222222222344222442222221
11111111222222222224444442222222
11111111222222222233443322222222
11111111112222222222222222222222
11111111112222222222222222222222
11111111111122222222222222222222
11221111111122222222222222222222
11332211111111112222222222222222
11332211111111112222222222222222
11111111111111112222222222222222
11111111111111112222222222222222
```

INPUT DATA

```
33333333333333333333111111111133
13333333333333333333311111111333
11333333333733333333331111113333
1113333337773333333333311133‾333
11113333333733333333333313333333
11111333333333333333333333333333
11111133333333333333333333333335
21111113333333333330033333333355
22111111133333333333333333333555
12211111133333333333333333555555
22111111143333331333333333555555
21111111144433333133333335555555
11111114444433333333333555555555
11111144444443333333333555555555
11114444444444411111111444444444
11114444444444441111114444444444
```

results from ADAPT1 , initial tolerance = 1.00

```
33333333333333333333111111111133
23333333333333333333322111112233
11333333355333333333331111113333
11223333557733333333332211333333
11113333333333333333333223333333
11112233333333333333333333333333
11111133333333333333333333333344
11111122333333333330033333333355
11111111133333333333333333334455
11111111133333333333333333345555
11111111133333333333333333555555
11111111133333333333333443555555
11111133444433333333333555555555
11111144444433333333334455555555
11113344444444111111111444444444
11114444444444433111134444444444
```

results from ADAPT1 , initial tolerance = .80

```
33333333333333333333111111111133
22333333333333333333322111112233
11333333355333333333331111113333
11223333557733333333332211333333
11113333333333333333333223333333
11112233333333333333333333333333
11111133333333333333333333333344
11111122333333333330033333333355
11111111133333333333333333334455
11111111133333333333333333445555
11111111133333333333333333555555
11111111133333333333333344555555
11111133444433333333333555555555
11111144444433333333334455555555
11113344444444111111111444444444
11114444444444431111334444444444
```

INPUT DATA


```
33333333333333333333111111111133
13333333333333333333111111111333
11333333337333333333311111113333
11133333777333333333331113333333
11113333337333333333333313333333
11111333333333333333333333333333
11111133333333333333333333333335
21111111333333333300333333333355
22111111333333333333333333333555
12211111133333333333333333355555
22111111143333331333333333555555
21111111444333331333333335555555
11111114444433333333333355555555
11111144444443333333333555555555
11111444444444111111144444444444
11114444444444411111144444444444
```

results from ADAPT1 , initial tolerance = .75

```
33333333333333333333111111111133
22333333333333333332211111112233
11333333553333333333311111113333
11223333557733333333332211333333
11113333333333333333333333333333
11112233333333333333333333333333
11111133333333333333333333333333
11111122333333333333333333333333
11111111333333333333333333334455
11111111333333333333333334455555
11111111333333333333333333555555
11111111333333333333333344555555
11111133444433333333333555555555
11111144444433333333334455555555
11113344444444411111144444444444
11114444444444331111334444444444
```

results from ADAPT1 , initial tolerance = .70

```
33333333333333333333111111111133
33333333333333333332211111112233
33333333333333333333311111113333
33333333333333333333322113333333
33333333333333333333333333333333
33333333333333333333333333333333
33333333333333333333333333333333
33333333333333333333333333333333
11111111333333333333333333334455
11111111333333333333333333445555
11111111333333333333333333555555
11111111333333333333333445555555
11111133444433333333333555555555
11111144444433333333334455555555
11113344444444411111144444444444
11114444444444331111334444444444
```

INPUT DATA


```
11112222222222444444444444444444
11111222222222244444444444444444
11111122222222224444444444444444
11111112222222222444444444444444
11111111222222222244444444444444
11111111122222222224444444444444
11111111112222222222444444444444
11111111111222222222244444444444
11111111111122222222224444444444
11111111111112222222222444444444
11111111111111222222222244444444
11111111111111122222222224444444
11111111111111112222222222444444
11111111111111111222222222244444
11111111111111111122222222224444
11111111111111111112222222222444
```

results from ADAPT2 , initial tolerance = .80

```
11112222222222224444444444444444
11112222222222224444444444444444
11112222222222224444444444444444
11112222222222224444444444444444
11111111222222224444444444444444
11111111222222224444444444444444
11111111222222224444444444444444
11111111222222224444444444444444
11111111222222222222224444444444
11111111222222222222222244444444
11111111222222222222222244444444
11111111222222222222222244444444
11111111111111112222222222444444
11111111111111111122222222224444
11111111111111111122222222224444
11111111111111111112222222224444
```

results from ADAPT2 , initial tolerance = .60

```
11112222222222224444444444444444
11112222222222224444444444444444
11111122222222224444444444444444
11111111222222224444444444444444
11111111222222224444444444444444
11111111222222224444444444444444
11111111222222224444444444444444
11111111222222224444444444444444
11111111111122222222224444444444
11111111111122222222222244444444
11111111111111222222222244444444
11111111111111112222222244444444
11111111111111112222222222444444
11111111111111111122222222224444
11111111111111111122222222224444
11111111111111111111222222224444
```

INPUT DATA

ORIGINAL PAGE IS
OF POOR QUALITY

```
11112222222222444444444444444444
11111222222222244444444444444444
11111122222222224444444444444444
11111112222222222444444444444444
11111111222222222244444444444444
11111111122222222224444444444444
11111111112222222222444444444444
11111111111222222222244444444444
11111111111122222222224444444444
11111111111112222222222444444444
11111111111111222222222244444444
11111111111111122222222224444444
11111111111111112222222222444444
11111111111111111222222222244444
11111111111111111122222222224444
11111111111111111112222222222444
```

results from ADAPT2 , initial tolerance = .40

```
11112222222222224444444444444444
11112222222222224444444444444444
11111122222222224444444444444444
11111111222222224444444444444444
11111111222222222244444444444444
11111111222222222222444444444444
11111111222222222222444444444444
11111111222222222222224444444444
11111111111122222222224444444444
11111111111122222222222444444444
11111111111111222222222244444444
11111111111111112222222244444444
11111111111111112222222224444444
11111111111111111122222222244444
11111111111111111122222222224444
11111111111111111111222222222244
```

results from ADAPT2 , initial tolerance = .10

```
11112222222222444444444444444444
11111122222222224444444444444444
11111122222222224444444444444444
11111111222222222244444444444444
11111111222222222244444444444444
11111111112222222222444444444444
11111111112222222222444444444444
11111111111122222222224444444444
11111111111122222222224444444444
11111111111111222222222244444444
11111111111111222222222244444444
11111111111111112222222222444444
11111111111111112222222222444444
11111111111111111122222222224444
11111111111111111122222222224444
11111111111111111111222222222244
```

APPENDIX

N83
23072

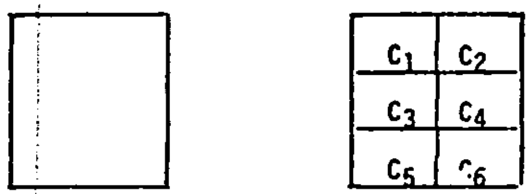
UNCLAS

This Page Intentionally Left Blank

FUNDAMENTAL RESEARCH DATA BASE

At the request of Dr. R. P. Heydorn, a fundamental research data base has been created on a single 9-track 1600 BPI tape containing ground truth, image, and Badhwar profile feature data for 17 North Dakota, South Dakota, and Minnesota agricultural sites. Each site is 5x6 nm in area. Image data has been provided for a minimum of four acquisition dates for each site. All four images have been registered to one another. A list of the order of the files on tape and the dates of acquisition is provided in attachment 1.

Attachment 2 provides information on the format of the ground truth tape and a table for each year to use in interpreting the information on the ground truth tape. Ground truth codes vary depending on the year. Like the Landsat image files, ground truth files cover an image 196 pixels wide by 117 lines long, but the actual size of the ground truth image is 392 pixels by 234 lines. The reason for this difference is that there are six ground truth subpixels for each Landsat pixel, as illustrated.



Landsat Pixel Ground Truth Pixel

The symbols C_1, C_2, C_3, C_4, C_5 and C_6 represent the ground truth crop code for the various sub-parts of the Landsat pixel. We typically use a plurality rule to decide on a single label for a Landsat pixel.

All files are stored on tape in universal format. Image files and Badhwar profile feature files contain four channels of data, but since three Badhwar profile features are provided in the feature files (t_p, σ, and G_{max}) the fourth channel is always zero. The format for image and profile files is the same and is provided in attachment 3.

ATTACHMENT 1

File	Type	Segment	Year	State	Acquisitions (Julian Date)						
1-5	Image	1380	78	MN	115	169	196	204	222		
6-10		1394	78	ND	120	174	211	220	238		
11-15		1531	77	MT	112	129	147	184	220		
16-20		1537	78		122	141	159	194	221		
21-27		1544	78	MT	104	122	140	158	176	221	230
28-32		1553	78	MT	122	194	203	211	220		
33-37		1566	78	MN	115	133	169	196	232		
38-43		1619	77	ND	122	140	158	175	176	230	
44-48		1636	78	ND	135	154	190	207	226		
49-53		1650	78	ND	156	191	209	218	236		
54-58		1653	78		136	154	155	191	208		
59-63		1663	77	ND	121	138	156	174	211		
64-68		1676	79	SD	120	165	184	211	237		
69-73		1755	79	SD	120	147	166	184	220		
74-78		1784	78	SD	133	169	196	223	241		
79-83		1825	78	MN	133	196	206	223	224		
84-88		1899	77	ND	122	140	157	175	193		
89-94	Image	1920	78	ND	101	136	199	209	217	236	

File	Type	Segment
95	GT	1380
96		1394
97		1531
98		1537
99		1544
100		1553
101		1566
102		1619
103		1636
104		1650
105		1653
106		1663
107		1676
108		1755
109		1784
110		1825
111		1899
112		1920

File	Type	Segment
113	Profile	1380
114		1394
115		1531
116		1537
117		1544
118		1553
119		1566
120		1619
121		1636
122		1650
123		1653
124		1663
125		1676
126		1755
127		1784
128		1825
129		1899
130		1920

Two end-of-files

ORIGINAL PAGE IS
OF POOR QUALITY

ATTACHMENT 2

3.2.1 HEADER RECORD

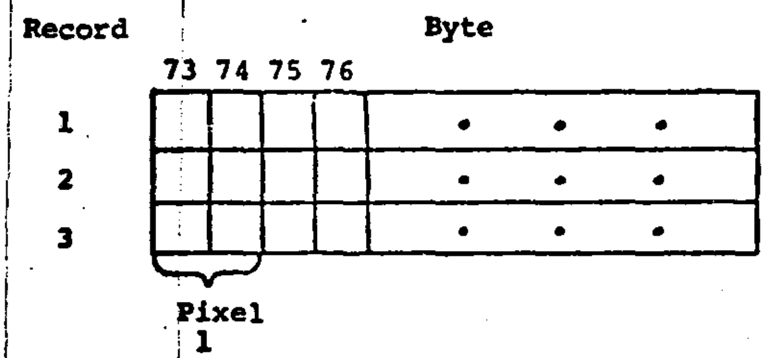
The Header Record is the first record on the tape and contains 3060 bytes (8 bits per byte). The record is zero filled except for those bytes listed in the following table. The values contained in the listed bytes are all constant except for bytes 61 through 63. The attached tape format contains identification and descriptions for each byte. The description and format of the Header Record is contained in attachment 3.

Byte	Value	Byte	Value	Byte	Value
61	Day	96	1	111	120
62	Month	97	-120	1778	1
63	Year	100	2	1786	1
81	-128	101	28	1787	1
89	1	104	1	1788	-120
90	1	106	70		
91	8	109	1		
93	1	110	1		

Each video scan line is 504 bytes long; a 2-byte record counter, a 70-byte ancillary block, and 392 bytes of ground truth (two of the six subpixels for a 196 pixel scan line). It takes three video scan lines to complete one scan line of ground truth. See the following page for diagram.

Each video block will be the same number of bytes in length. If this tape contains raw data the PCM sync words associated with the video data, if any, will be included, with the video data on this tape. If this tape contains processed data, no sync words will be present.

The arrangement of data for each pixel is shown in the following diagram. Data for subpixels 1 and 2 for pixel 1 is found in bytes 73 and 74 of the first data record. Data record 2 and 3 contain data for suppixel 3 through 6 in the same format as record 1.



ORIGINAL PAGE IS
OF POOR QUALITY

C Sorensen.

APPROVED SYMBOL LIST

1976 & 1977

SYMBOL	DESCRIPTION	GREY SCALE LEVEL	HARVESTED	ABANDONED	STRIP FALLOW	STRIP FALLOW ABANDONED	STRIP FALLOW HARVESTED
A	ALFALFA	90	115	140	165	190	215
B	BARLEY	101	126	151	176	201	226
BN	BEANS	91	116	141	166	191	216
C	CORN	92	117	142	167	192	217
CN	COTTON	111	136	161	186	211	236
FX	FLAX	103	128	153	178	203	228
G	GRASS	105	130	155	180	205	230
H	HAY	106	131	156	181	206	231
I/CC	IDLE COVER CROP	252	-	-	-	-	-
I/CS	IDLE CROPLAND STUBBLE	251	-	-	-	-	-
I/F	IDLE CROPLAND FALLOW	254	-	-	-	-	-
I/RE	IDLE CROPLAND RESIDUE	253	-	-	-	-	-
M	MILLET	112	137	162	187	212	237
MT	MOUNTAINS	241	-	-	-	-	-
NA	NON-AG	242	-	-	-	-	-
O	OATS	104	129	154	179	204	229
P	PASTURE	107	132	157	183	207	232
PF	PROBLEM FIELD	80	-	-	-	-	-
R	RYE	102	127	152	177	202	227
SB	SUGAR BEETS	98	123	148	173	198	223
SF	SAFFLOWER	93	118	143	168	193	218
SG	SUDAN GRASS	95	120	145	170	195	220
SR	SORGHUM	96	121	146	170	196	221
SU	SUNFLOWER	94	119	144	169	194	219
SW	SPRING WHEAT	100	125	150	175	200	225
SY	SOYBEANS	97	122	147	172	197	222
T	TREES	108	133	158	183	208	233
TR	TRETRICALE	109	134	159	184	209	234
VW	VOLUNTARY WHEAT	110	135	160	185	210	235
W	WINTER WHEAT	99	124	149	174	199	224
*	WATER	240	-	-	-	-	-
X	HOMESTEAD	250	-	-	-	-	-

178

TY CROP CODE 1978 ATTACH 2

•Changes from 1977-1978 Codes (1-30)

			HARVESTED	ABANDONED	STRIP FALLOW	STRIP FALLOW ABANDONED	STRIP FALLOW HARVESTED
20	Problem Field	(PF)					
90	Alfalfa	(A)	115	140	165	190	215
91	Beans	(BN)	116	141	166	191	216
92	Corn	(C)	117	142	167	192	217
93	Safflower	(SF)	118	143	168	193	218
94	Sunflower	(SU)	119	144	169	194	219
• 95	Durum Wheat	(DW)	120	145	170	195	220
96	Sorghum	(SR)	121	146	171	196	221
97	Soybeans	(SY)	122	147	172	197	222
98	Sugar Beets	(SB)	123	148	173	198	223
99	Winter Wheat	(WW)	124	149	174	199	224
100	Spring Wheat	(SW)	125	150	175	200	225
101	Spring Barley	(BS)	126	151	176	201	226
102	Rye	(R)	127	152	177	202	227
103	Flax	(FX)	128	153	178	203	228
104	Spring Oats	(SO)	129	154	179	204	229
• 105	Fall Oats	(FO)	130	155	180	205	230
• 106	Fall Barley	(FB)	131	156	181	206	231
• 107	Cotton	(CN)	132	157	182	207	232
108	* Peanuts	(PN)	133	158	183	208	233
109	*		134	159	184	209	234
110	*		135	160	185	210	235

ORIGINAL PAGE IS OF POOR QUALITY

• 111 Grass (G)

• 112 Hay (H), Sudan Grass (SG), Millet (ML) *Open - to be assigned as needed

• 113 Pasture (P)

• 114	Trees	(T)	136	* Pasture Mix (PM)	188	* CL (cover) 188	243	[illegible]
			137	* WW to SW Grain	189	*	244	•
140	Water	(*)	138	* WW to Annual Crops	211	*	245	•

Attachment 2 to Ref: 642-7665
1979 Crop Year Keys and Delineation Codes

Crop Type	Crop Key	Crop Harvested	Crop Abandoned	Crop Harvested for Silage
Alfalfa	AH	101	151	201
Buckwheat	BW	102	152	202
Barley	BK	103	153	203
Clover	CL	104	154	204
Corn	CR	105	155	205
Cotton	CT	106	156	206
Dry Bean	DB	107	157	207
Durum Wheat	DW	108	158	208
Flax	FX	109	159	209
Millet	ML	110	160	210
Oats	OA	111	161	211
Peanuts	PE	112	162	212
Potatoes	PO	113	163	213
Rice	RI	114	164	214
Rye	RY	115	165	215
Sugar Beets	SB	116	166	216
Sugar Cane	SC	117	167	217
Safflower	SF	118	168	218
Soybeans	SO	119	169	219
Sorghum	SR	120	170	220
Sunflower	SU	121	171	221
Spring Wheat	SW	122	172	222
Tobacco	TB	123	173	223
Vegetables	VE	124	174	224
Winter Wheat	WW	125	175	225
Small Grains/Strip Fields	--	126	176	226
*		127	177	227
*		128	178	228

Grasses	GS	131		
Other Hay	OH	132		
Orchard/Vineyards	OR	133		
Pasture	PA	134		
Trees	TR	135		
Water > 5 acres	WA	136		
Non-Agriculture	XX	140		
Idle Land/Fallow	IL	231		
Previous Year Residue/Stubble	R	232		
Mixed Crop in Field	M	233		
Problem Field		99		
Non-Inventoried Land		255		

*Open--to be assigned as needed (through code 130). Other open codes include 137 through 139, 141 through 150, 181 through 200, 234 through 254.

ATTACHMENT 3

ORIGINAL PAGE IS
OF POOR QUALITY

UNIVERSAL FORMAT TAPE HEADER RECORD FORMAT (3060 Bytes)

BYTE	CONTENTS	DESCRIPTION
1-32	LACIE ѢNDPFѢ...Ѣ	Computing system id-EBCDIC
*33-38	XXXXXX	6-digit unload tape number
*39-52	PYYDDDHHMMSSTH	RUNID (EBCDIC)
53-60	ERTSѢMSSѢ...Ѣ	Sensor id-EBCDIC
61-63		Date of this tape generation
61		Day of month - Binary
62		Month number - Binary
63		Year - last 2 digits - Binary
64	8	Daily tape serial number - Binary
65-66		ERTS mission number - Binary 1 = ERTS A 2 = ERTS B
67-68		Site - Binary (sample segment number) Range 1-5000
69	00000000	Line - Binary
70	00000000	Run - Binary
71-72		Orbit number of new data - Binary
73-80		Time of first scan in this job (for LACIE this is the time of the center scan of the ERTS scene containing the sample segment to the last ten seconds)
73-74		Tenths of seconds x 1000 - Binary
75		Seconds - Binary
76		Minutes - Binary
77		Hours - Binary
78		Day of month - Binary
79		Month number - Binary
80		Year - last 2 digits - Binary
81-88		Bands active in this job, 1 bit per band left to right (MSB to LSB). Video data always appears in the order indicated here. 1 = active.
81	11110000	Bands 1, 2, 3, 4 active
82-88	0	Bands 5-64 not applicable to LACIE
89	0	Processing flag - raw data - Binary
90	4	Number of bands in this job - Binary

BYTE	CONTENTS	DESCRIPTION
91	8	Number of bits in a picture element - Bin
92-93	1	Address of start of video data gives loca of start of video within scan - Binary
94-95	0	Address of start of first calibration area within the scan - Binary
96-97	196	Number of video elements per scan within a single band - Binary
98-99	0	Number of calibration elements in the first calibration area within the scan in a single band - Binary
100-101	900	Physical record size in bytes - Binary
102	0	Number of bands per physical record of data set starting with the second record of the data set - Binary
103	0	Number of physical records per scan per band - Binary. Zero unless the elements per band is greater than 3K.
104	1	Number of records to make a complete data set - Binary.
105-106	70	Length of ancillary block in bytes - Binary
107	0	Data order indicator - Binary 0 = video ordered by band
108-109	1	Start pixel number number of the first pixel per scan on this tape referenced to the start of the scan - Binary
110-111	196	Stop pixel number number of the last pixel per scan on this tape referenced to the start of the scan - Binary
112-623		Coefficients and exponents-of-ten to linearly translate parameter values from up to 64 bands to engineering units. Two bytes per coefficient or exponent with each pair of bytes expressed in signed binary. (MSB a sign bit: 0=+, 1=-. (Remaining 15 bits straight binary).
112-119	0	A0 coefficients for bands 1-4
120-239	0	Bands 5-64 not applicable to LACIP
240-247	0	E0 exponents of ten for bands 1-4
248-367	0	Bands 5-64 not applicable to LACIE
368-369	1	A1 coefficient for band 1
370-371	1	A1 coefficient for band 2

BYTE	CONTENTS	DESCRIPTION
371-373	1	A1 coefficient for band 3
374-375	1	A1 coefficient for band 4
376-495	0	Bands 5-64 not applicable to LACIE
496-503	0	E1 exponents of ten for bands 1-4
504-629	0	Bands 5-64 not applicable to LACIE where for each band Y = Engineering Units, C = Parameter Value: Y = A *10**E + C*A *10**E
624-687	To be supplied by JSC	Color code information - one byte per band in same order as "channel active on this tape" indicator - Binary. 0 = no color assignment
688-751	0	Scale factor - one byte per band in same order as "channel active on this tape" indicator - Binary 0 = not active
752	0	Offset constant - Binary
753	16	Word size of generating computer. This is the smallest quantity in bits that the computer can write on tape.
754-1777		Shortest and longest wave-length of each band - EBCDIC. Eight bytes per limit, 16 bytes per band - mili microns
754-769	0000050000000600	Band 1 - EBCDIC
770-785	0000060000000700	Band 2 - EBCDIC
786-801	0000070000000800	Band 3 - EBCDIC
802-817	0000080000001100	Band 4 - EBCDIC
818-1777	0	Bands 5-64 not applicable to LACIP - EBCDIC
1778	1	Number of data sets per physical record - Binary
1779-1780	0	Address of start of second calibration within a scan - Binary
1781-1782	0	Number of calibration elements in the second calibration area within the scan in a single band - Binary
1783	0	Calibration source indicator - Binary
1784	0	Fill zero.
1785-1786	4	Number of bands in the first record of the data set - Binary
1787-1788	196	Total number of elements per scan per band - Binary

ORIGINAL PAGE IS OF POOR QUALITY

BYTE	CONTENTS	DESCRIPTION
1789-1790	1	Pixel skip factor - the quantity to be to the number of the last pixel process yield the number of the next pixel to be processed - Binary 1 = Process every pixel
1791-1791	1	Scan skip factor - the quantity to be add to the number of the last scan processed yield the number of the next scan to be processed - Binary. 1 = Process every scan
1793-2940		General information. Information in EBCDIC generated to satisfy user requirements. Contents will be unique for each user and depend not only on the sensor, but also on specifications of the user for whom the tape is generated. Bytes for which user specific no requirements will contain fill zeros.
1793-2086		Fill zeros
2087-2184		General annotation byte assignment for ERTS ZACIE
2087-2094	±X.XXXXX	Peak sharpness - EBCDIC
2095-2102	±X.XXXXX	Normalized peak to background ratio - EBCDIC
2103		Manual registration flag 0 = Automatic 1 = Manually assisted
2104		Zero fill flag - Binary 0 = The sample segment contains no zero fill data 1 = Part of the sample segment contains zero fill data
2105-2106		Orbit number of reference data set - Binary (not used = 0)
2107-2109		Zero fill
2110		Cloud cover - Binary - percent of 10X11 KM search area covered by clouds
2111		Zero fill
2112-2120		ERTS scene/frame id number for reference data set - EBCDIC - ADDDHHMMS (see bytes 2123-2131 for content)
2121		Zero fill
2122		Flag indicating whether a reference scene has been used for registration - Binary 0 = hasn't been used 1 = has been used
2123-2131		ERTS scene-frame id number for new data-EBCDIC-ADDDHHMMS

ORIGINAL PAGE IS OF POOR QUALITY

BYTE	CONTENTS	DESCRIPTION
2123		A = ERTS mission number
2124-2126		DDD = Day number relative to launch at time of observation
2127-2128		HH = hour at time of observation
2129-2130		MM = minute at time of observation
2131		S = tens of seconds at time of observation
2132		Zero fill
2133		Data quality classification 0 = acceptable 1 = marginal
2134-2145		Center of sample segment - EBCDIC right justified and padded with zeros
2134-2139		Latitude
2134		"N" = North "S" = South
2135-2137		Degrees - integral
2138-2139		Minutes - integral
2140-2145		Longitude
2140		"E" = East; "W" = West
2141-2143		Degrees - integral
2144-2145		Minutes - integral
2146-2149		Band sync status - Binary - the number of lines for which sync could not be maintained during pre-processing by band
2146		Band 1
2147		Band 2
2148		Band 3
2149		Band 4
2150-2156		Zero fill
2157-2170		Sun angle - EBCDIC
2157-2162	SUN℔EL	"SUN EL" - EBCDIC
2163-2164		Sun elevation - integral degrees EBCDIC
2165-2167	℔AZ	"AZ" - EBCDIC
2168-2170		Sun azimuth - integral degrees - EBCDIC
2171-2178		Time and date of last update to controlling information - EBCDIC - YDDDHHMM
2179-2184		Zero fill

BYTE	CONTENTS	DESCRIPTION
*		Sun angles are 2 byte binary
*2201-2202		Sun angle for RSEG channels 1-4
*2203-2204		Sun angle for RSEG channels 5-8
*2205-2206		Sun angle for RSEG channels 9-12
*2207-2208		Sun angle for RSEG channels 13-16
*2249	YYDDD	1st acquisition date (characters)
*2254	X	Average soil greenness for 1st acquisition (binary number)
*2257	YYDDD	2nd acquisition date or blanks
*2262	X	Average soil greenness for 2nd acquisition
*2265	YYDDD	3rd acquisition date or blanks
*2270	X	Average soil greenness for 3rd acquisition
*2273	YYDDD	4th acquisition date or blanks
*2278	X	Average soil greenness for 4th acquisition
2551-2642	0	General annotation byte assignments for the cyber at JSC
2643-2940		General annotation byte assignments for the production film converter
2643-2658		Bias factors and scaling factors - signed Binary. Four bytes per channel, where first two bytes = bias factor; second two bytes = scaling factor. Each factor has an implied decimal point to the left of the least significant decimal digit. If MSB = 1 the factor is negative; if the MSB = 0 the factor is positive.
*2643-2646		Channel 1
*2643-2644		Bias factor
*2645-2646		Scaling factor
*2647-2650		Channel 2
*2647-2648		Bias factor
*2649-2650		Scaling factor
*2651-2654		Channel 3
*2651-2652		Bias factor
*2653-2654		Scaling factor
*2655-2658		Channel 4
*2655-2656		Bias factor
*2657-2658		Scaling factor

ORIGINAL PAGE IS OF POOR QUALITY

BYTE	CONTENTS	DESCRIPTION
*2659-2606		Bias factor and scaling factors for channels 5-16 in the same format as above.
2758		
2759	1	N thousand scan lines per frame - Binary
*2760-2783		User ID
*2784-2789		Blanks
2790-2792	0	Altitude in meters - Binary
2793-2794	0	Ground speed in MET/SEC - Binary
2795	1	Scan Type - Binary 00000000 = Raw data 00000001 = Smoothed data
2796	0	Angle of ABC in degrees - Binary
2797	1	Camera - Binary 00000000 = 70 MM 00000001 = 5 inch
2798	0	Input device - Binary 00000000 = 9-track 00000001 = high density tape
2799	2	Truncation 0 = 2 low order bits 1 = 2 high order bits 2 = no truncations
2800-2807		Channels requested. 1 bit per channel - Binary
2800-2801	11110000 00000000 (1 acq) 11111111 00000000 (2 acq) 11111111 11110000 (3 acq) 11111111 11111111 (4 acq)	Channels 1, 2, 3, 4 requested
2802-2807	0	Channels 16-64 not applicable for Unload
2808	0	Processing mode - Binary 00000000 = serially 00000001 = concurrently
2809-2824	0	Density for eight saturated colors - two bytes per saturated color - Binary where first byte = low intensity level of the range; second byte = high the range of the intensity level is 0 to 255
2809-2810		Red density range
2811-2812		Blue density range
2813-2814		Green density range
2815-2816		Magenta density range
2817-2818		Cyan density range
2819-2820		Yellow density range

ORIGINAL PAGE IS OF POOR QUALITY

BYTE	CONTENTS	DESCRIPTION
2821-2822		White density range
2823-2824		Black density range
2825	To be supplied by JSC	Film processing flag 0 = Process this file 1 = Skip this file
2826-2873	0	Fill zero
2874	0	Color select* - Binary 0 = No color 1 = Assigned color 2 = False color 3 = Saturated color
2875	0	Image format* - Binary 0 = Single image 1 = Enhanced images 2 = Abut images 3 = Offset images
2876	6	Repeat of pixels per scan - Binary 0 = None 1 = 1 repeat 2 = 2 repeats n = n repeats
2877	8	Repeat of scan - Binary 0 = none 1 = 1 repeat 2 = 2 repeats n = n repeats
2878-2881		Partial scan - Binary
2878-2879	0	Start pixel number
2880-2881	0	Stop pixel number (If bytes 2787-2861 contain all zeros, full scan is expected - not partial)
2882-2883	0	Sensor scan rate in scans/second - Binary
2884	0	Pixel size - Binary
2885-2886	0	Angle of drift - Binary
2885		± integer degrees
2886		Fraction
2887-2940	0	Fill zeros
2941-3000	LACIEb̸NDPFb̸...b̸	Title - user designated identification
3001-3060	0	Fill zeros, makes the record an integral number of computer words. These bytes must <u>never</u> contain data.

3.2.2 DATA SETS

The data follows the Header Record and is arranged in data sets. A data set is defined as the ancillary data and all of the video data for one scan line for all active channels. Data sets are recorded in variable length physical records containing a maximum of 3000 bytes of information per record. Since 3000 bytes is not compatible with the word length of all computers, the record includes a sufficient number of fill zero to make the record divisible by 32, 36, 48, and 60 bits. However, the maximum length of the record may not exceed 3060 bytes. If two are more records are needed for the data set, the data set will be divided. Under no condition will the data for a video channel begin in one record and continue into another record.

The first two bytes of each record will contain the number of the physical record within the video data set. This is for use in data sets that contain more than 3000 bytes and therefore require more than one physical record for recording. The ancillary block is the first block of a data set and follows the record counter. The length of the ancillary block is variable, with the number of bytes given in the header record.

Bytes 73 through N will be dependent on whether this job contains raw processed data (Byte 89 of the header record). The value of N will be given in bytes 105 and 106 of the header record and will always be greater than or equal to 70.

If this job contains raw data bytes 73 through N will contain the housekeeping data channel from the sensor, if one is available.

Following the ancillary data in each data set will be the video data for the one channel for one scan. The video data for the channel for one scan will comprise a video block.

UNIVERSAL FORMAT TAPE ANCILLARY BLOCK FORMAT

BYTE	CONTENTS	DESCRIPTION
1-68	0	Zero fill
69-70		Relative scan line number

UNIVERSAL FORMAT SCAN LINE FORMAT (900 Bytes)

ORIGINAL PAGE IS
OF POOR QUALITY

- HEADER 1060 bytes → A
- IRG
- DATA SET 1 Variable number of bytes → B
- IRG
- ⋮ (DATA RUN 1)
- IRG
- DATA SET 6 (line 6) Variable length
- IRG
- ⋮
- IRG
- DATA SET 117 (line 117)
- EOF
- HEADER (DATA RUN 2)
- EOF
- HEADER
- EOF
- EOF
- EOF

ONE SAMPLE SEGMENT

Note: The number of bytes for each data set will be the same in each data run.

A and B are retrieved from the image data base.
A= LSIMAGHD , master header
B= LSIMVHDR, LSIMCHAN, LSIMLAHD, imagery data

(B$_1$)

RECORD COUNTER		2 bytes
ANCILLARY BLOCK		70 bytes
LINE 1,	BAND 1	196 bytes
	BAND 2	196 bytes
	BAND 3	196 bytes
	BAND 4	196 bytes
ZERO FILL		44 bytes

DATA SET FOR 1 ACQUISITION, 4 CHANNELS

900 BYTES/RECORD

Figure 3-15. PFC Unload Tape (sheet 1 of 2)

(B_3)

RECORD COUNTER		2 bytes
ANCILLARY BLOCK		70 bytes
LINE 1,	BAND 1	196 bytes
	BAND 2	196 bytes
	BAND 3	196 bytes
	BAND 4	196 bytes
	BAND 5	196 bytes
	BAND 6	196 bytes
	BAND 7	196 bytes
	BAND 8	196 bytes
	BAND 9	196 bytes
	BAND 10	196 bytes
	BAND 11	196 bytes
	BAND 12	196 bytes
ZERO FILL		96 bytes

DATA SET FOR THREE ACQUISITION,
12 CHANNELS

2520 BYTES/RECORD

Note: For a 16-channel data set, two (B_2) data sets will be required therefore requiring two physical records.

Figure 3-15. PFC Unload Tape (Sheet 2)

END

DATE

FILMED

JUN 30 1983

End of Document