https://ntrs.nasa.gov/search.jsp?R=19780006514 2020-03-22T05:47:09+00:00Z

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

N78-14457

(E78-10040) SIGNATURE EXTENSION PREPROCESSING FOR LANDSAT MSS DATA Final Report, 15 May 1976 - 14 Nov. 1977 (Environmental Research Inst. of Michigan) 69 p HC A04/MF A01 CSCL 05B G3/43 Unclas 00040

E78-10040

NASA CR- 151563
ERIM 122700-32-F

"Made available under NASA sponsorship In the interest of early and wide dissemination of Earth Resources Survey Program information and without liability for any use made thereof."

Final Report

SIGNATURE EXTENSION PREPROCESSING FOR LANDSAT MSS DATA

PETER F. LAMBECK
Infrared and Optics Division

NOVEMBER 1977

Principal Investigator
Richard F. Nalepka

DEC 1977 RECEIVED NASA STI FACILITY

Prepared for
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Johnson Space Center
Earth Observations Division
Houston, Texas 77058
Contract No. NAS9-14988, Task 2
Technical Monitor: I. Dale Browne/SF3

ENVIRONMENTAL
RESEARCH INSTITUTE OF MICHIGAN
FORMERLY WILLOW RUN LABORATORIES, THE UNIVERSITY OF MICHIGAN
BOX 8618 • ANN ARBOR • MICHIGAN 48107

JAN 1978 RECEIVED NASA STI FACILITY

ORIGINAL PAGE IS
OF POOR QUALITY

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
122700-32-F		

4. Title and Subtitle	5. Report Date
Signature Extension Preprocessing for Landsat MSS Data	November 1977
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
P. F. Lambeck	122700-32-F
	10. Work Unit No.
	Task 2

9. Performing Organization Name and Address	11. Contract or Grant No.
Environmental Research Institute of Michigan Infrared & Optics Division P.O. Box 8618 Ann Arbor, Michigan 48107	NAS9-14988
	13. Type of Report and Period Covered Final Technical Report 15 May 76 - 14 Nov 77
12. Sponsoring Agency Name and Address National Aeronautics & Space Administration Johnson Space Center Houston, Texas 77058	14. Sponsoring Agency Code

15. Supplementary Notes

The work was performed for the Earth Observations Division.
I. Dale Browne (SP3) was the Technical Monitor.
The ERIM Principal Investigator on this contract was Richard F. Nalepka

16. Abstract

Signature extension is a process intended to increase the spatial-temporal range over which a set of training statistics can be used to classify data without significant loss of recognition accuracy. This process is intended to help minimize the requirements for collecting ground truth and for extracting training statistics, thus allowing more timely and cost-effective surveys over large land areas. The reported effort has been primarily focussed to aid in performing large area agricultural surveys, using data from the Landsat satellites.

Current signature extension preprocessing techniques which have been developed or investigated at ERIM are presented. The discussion covers the underlying theory for the preprocessing, the development of haze correction algorithms (specifically XSTAR and XBAR), the development of an automatic screening procedure (SCREEN) to detect garbled data, clouds, snow, cloud shadows, and water in Landsat MSS data, results from tests of the preprocessing performance, some analyses of soil color effects in Landsat data, and conclusions and recommendations for future developments in preprocessing.

The results presented indicate significant success in preprocessing Landsat agricultural data to compensate for the effects of atmospheric haze without relying on ground observations, and also show promise for further significant improvements in the near future.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Remote Sensing Multispectral Processing Signature Extension Landsat Haze Correction	Initial distribution is listed at the end of this document.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified		

*For sale by the National Technical Information Service, Springfield, Virginia 22161

Original photography may be purchased from:
EROS Data Center
Sioux Falls, SD

PREFACE

This report describes part of a comprehensive and continuing program of research in multispectral remote sensing of the environment from aircraft and satellites and the supporting effort of ground-based researchers in recording, coordinating, and analyzing the data gathered by these means. The basic objective of this program is to improve the utility of remote sensing as a tool for providing decision makers with timely and economical information from large geographical areas.

The feasibility of using remote sensing techniques to detect and discriminate between objects or conditions at or near the surface of the earth has been demonstrated. Applications in agriculture, urban planning, water quality control, forest management, and other areas have been developed. The thrust of this program is directed toward the development and improvement of advanced remote sensing systems and includes assisting in data collection, processing and analysis, and ground truth verification.

The research covered in this report was performed under NASA Contract NAS9-14988. The program was directed by R. R. Legault, Director of ERIM's Infrared and Optics Division and an Institute Vice-President, Q. A Holmes, Head of the Information Systems and Analysis Department and Project Director, and R. F. Nalepka, Head of the Multispectral Analysis Section (MAS) and Principal Investigator. The Institute number for this report is 122700-32-F.

The author wishes to acknowledge the administrative direction provided by Mr. R. R. Legault, Dr. Q. A. Holmes, and Mr. R. F. Nalepka and the technical assistance given by Mr. R. F. Nalepka, Dr. W. A. Malila, Mr. R. J. Kauth, Mr. J. F. Hemdal, Mr. J. K. More, and Dr. R. E. Turner. Ms. D. Dickerson, E. Hugg, and M. Warren are thanked for their secretarial assistance.

CONTENTS

FIGURES

FIGURES (Cont'd)

TABLES

SUMMARY

The general form of the transfer equation representing the recorded MSS signal level in each spectral band for a given material indicates that differences in recording conditions between a training scene and a recognition scene cause multiplicative and additive changes in the signal levels observed. Although the effects of bidirectional reflectance can cause these multiplicative and additive changes to be unique for each material, generalized multiplicative and additive data transformations can be derived which provide significant compensation for differences in recording conditions between training and recognition scenes.

Previous investigations of signature extension techniques relying on a dependable correlation between the statistical data distributions for training and recognition areas (e.g., cluster matching algorithms) have indicated that such procedures are at present unreliable due to the unpredictable frequent occurrence of significant differences in training and recognition scene composition. Subsequent signature extension efforts at ERIM have attempted to circumvent this difficulty principally by focusing attention on preprocessing techniques which compensate only for identifiable physical effects (haze, viewing and illumination geometry) and devising methods (multisegment or multitemporal training) for extracting more completely representative training statistics. This report summarizes our progress in developing preprocessing techniques to compensate Landsat MSS data for physical effects without using ground observations.

Two signature extension preprocessing algorithms, XSTAR and XBAR, have been developed. The XSTAR algorithm is an uncomplicated technique which has been shown to provide significant and reliable compensation for the effects of atmospheric haze and sun illumination angle in Landsat

agricultural MSS data. The XBAR algorithm, currently under development as an improvement upon XSTAR, is a more sophisticated technique designed to provide compensation for the effects of atmospheric haze, sun illumination, view angle, and background albedo. The XBAR algorithm is based on detailed use of the ERIM radiative transfer model.

A data screening step to identify and eliminate confusing information within a scene, such as garbled data, clouds, snow, cloud shadows, and water, is necessary prior to calculating the haze diagnostics needed by the XSTAR and XBAR algorithms. A fully automatic screening procedure (called SCREEN) for Landsat MSS data has been developed for this purpose. The output from SCREEN is generally accurate enough to be used to edit the input to a classifier, however better results can be obtained through data analyst interaction with the SCREEN output.

Some analyses have been performed to estimate the effect of soil color on Landsat signals from agricultural areas, however these analyses have been hampered by a lack of adequate ground truth information during portions of the growing season when soils are distinguishable. Through signature modeling and analysis of the limited Landsat data with ground truth that is available, some of the variation caused by soils has been characterized.

Current progress in preprocessing for signature extension indicates that although some significant gains have been made with the XSTAR and SCREEN algorithms, an additional reduction by a factor of 2 in the signal differences between Landsat scenes should be possible in the near future. There is also a need to begin developing similar techniques for other sensors (e.g., the Thematic Mapper) and to test the present techniques in non-agricultural applications.

INTRODUCTION

Signature extension is a process intended to increase the spatial-temporal range over which a set of training statistics can be used to classify data without significant loss of recognition accuracy. The training statistics which are required are extracted from multispectral scanner (MSS) data with the aid of training information (ground truth) obtained from localized surveys on the ground or from interpretation of aerial photographs or MSS data images by trained analyst interpreters (AI's). Either of these procedures for acquiring ground truth information becomes costly and time consuming even for data processing over land areas of moderate size.

The goal of signature extension is to minimize the requirements for collecting ground truth and for extracting training statistics, thus reducing the associated costs and time delays. Signature extension would then help to provide timely and cost-effective classification over extensive land areas, including remote areas for which ground truth information may not be readily available. ERIM's present signature extension effort has been primarily concerned with the problem of performing large area agricultural surveys, using MSS data from the Landsat satellites.

Previous investigations of signature extension techniques relying on a dependable correlation between the statistical data distributions for training and recognition areas (e.g., cluster matching algorithms [1]) have indicated that such procedures are at present unreliable due to the unpredictable frequent occurrence of significant differences in training and recognition scene composition. Subsequent signature extension efforts at ERIM have attempted to circumvent this difficulty principally by focusing attention on preprocessing techniques which compensate only for identifiable physical effects (haze, viewing and

illumination geometry) and devising methods (multisegment or multitemporal training) for extracting more completely representative training statistics. This report summarizes our progress in developing preprocessing techniques to compensate Landsat agricultural MSS data for physical effects without using ground observations. Specific topics which are discussed include:

1. The underlying theory for physical effects compensations
2. The SCREEN procedure for automatically detecting garbled data, clouds, snow, cloud shadows, and water in Landsat MSS data
3. The XSTAR signature extension preprocessing algorithm
4. The XBAR signature extension preprocessing algorithm
5. Analyses of the effects of soil color or soil conditions on agricultural Landsat data.

Current progress at ERIM in other related aspects of the signature extension problem is reported in References 2, 3, and 4.

3
THEORY

3.1 EXPECTATIONS

The general form of the transfer equation representing the recorded MSS signal level in each spectral band for a given material indicates that differences in recording conditions between a training scene and a recognition scene cause multiplicative and additive changes in the signal levels observed. Although the effects of bidirectional reflectance can cause these multiplicative and additive changes to be unique for each material, generalized multiplicative and additive data transformations can be derived, based upon identifiable physical effects, which provide significant compensation for differences in recording conditions between training and recognition scenes. This will be demonstrated in the sections which follow.

Successful preprocessing techniques compensating for physical effects in Landsat data can provide several benefits, for example:

1. Allow training statistics to be derived from more than one region within a partition to provide more complete and representative training information
2. Remove the need for cluster matching algorithms, which are prone to failure whenever the scenes compared are not nearly equivalent subsets of the data distribution to be expected within a partition
3. Provide a stable data base from which to identify distinct crop growth trends to be used to identify crop types in unitemporal or multitemporal data.

The development of such techniques starts with a basic understanding of how physical factors affect the recorded signals from the scanner.

3.2 THE RADIATIVE TRANSFER EQUATION [5,6]

The radiance, L, at a given wavelength, observed by a satellite while viewing a target with reflectance ρ_t is represented by

$$L = \tilde{E}_{-}(\tau) \frac{\rho_t}{\pi} e^{-\tau/\mu} + L_p \tag{1}$$

with $\tilde{E}_{-}(\tau)$ representing the sum of the direct and diffuse irradiance on the target, τ representing the optical thickness of the atmosphere (denoted by τ_o in Reference 5), μ representing the cosine of the viewing angle relative to nadir, and L_p representing the path radiance due to scattering in the atmosphere. According to ERIM's radiative transfer model [5], Equation 1 may be expanded (and rearranged) as

$$L = \Lambda \left[\frac{4\mu_o^2 \rho}{\Delta} \left(\left\{ C_o \frac{\rho_t}{\rho} + \frac{1}{2} \frac{\tau}{\mu} \right\} e^{-\tau/\mu} + C_1 \left\{ 1 - \left[1 + \frac{\tau}{\mu} \right] e^{-\tau/\mu} \right\} \right) \right.$$

$$\left. + C_2 \left\{ \left[1 - \frac{\tau}{\mu} \right] - e^{-\tau/\mu} \right\} + \frac{\tau}{\mu} \right] \tag{2}$$

with

$$\Lambda = \frac{E_o}{4\pi} [\mu_o + (1-\eta)\tau]^{-1} \tag{3}$$

$$\Delta = 1 + 2(1-\rho)(1-\eta)\tau \tag{4}$$

$$C_o = 1 + 2(1-\eta)\tau \tag{5}$$

$$C_1 = \frac{1}{2} + 2(1-\eta)\mu \tag{6}$$

$$C_2 = [\mu_o - (1-\eta)\mu]\, p(\mu,\phi,-\mu_o,\phi_o) - [(1-\eta)\mu]\, p(\mu,\phi,\mu_o,\pi+\phi_o) \quad (7)$$

$$\eta = \frac{0.5\,\tau_R + 0.95\,\tau_A}{\tau_R + \tau_A} \quad (8)$$

$$\tau = \tau_R + \tau_A \quad (9)$$

Equations 2 through 9 approximate the effects of an atmosphere without absorption. The appropriate equations for an atmosphere with absorption are given in Reference 6, and can be arranged into an algebraic form analogous to Equations 2 through 9, however, for the present, atmospheric absorption will not be treated in this discussion. All of the variables in these equations are functions of wavelength with the exception of the geometric parameters μ, ϕ, μ_o, and ϕ_o, which represent the cosine of the viewing angle relative to nadir, the viewing azimuth, the cosine of the solar zenith angle, and the solar azimuth, respectively. The functions $p(\mu,\phi,-\mu_o,\phi_o)$ and $p(\mu,\phi,\mu_o,\pi+\phi_o)$ represent scattering phase functions. The anisotropy parameter, η, represents the fraction of scattered radiation which is scattered into the forward hemisphere, and is a weighted average of the anisotropy for Rayleigh scattering and for aerosol scattering (Equation 8). The optical thickness, τ (called τ_o in Reference 5), is the sum of the Rayleigh optical thickness, τ_R (which is small and does not vary with changes in the atmospheric state), and the aerosol optical thickness, τ_A (which is typically from three to twenty times larger than τ_R). The background albedo, ρ, is the average reflectance of the scene surrounding the target. The direct solar irradiance at the top of the atmosphere is represented by E_o.

The quantities Λ, Δ, C_o, C_1, and C_2 are all weak functions of τ (and ρ), varying by at most $\pm 5\%$ for reasonable atmospheric conditions,

except when the sun zenith angle is high (e.g., $\sim 70^{\circ}$). In the latter case A and C_2 may vary by up to $\pm 10\%$ with changes in τ. Thus, the major dependence of the radiance, L, on the optical thickness, τ, (and the background albedo, ρ) is shown explicitly in Equation 2. Note, however, that the quantity C_2 is a strong function of viewing angle, varying by $\pm 87\%$ at a sun zenith angle of 30°, when the viewing angle varies by $\pm 5.5^{\circ}$. This sensitivity of the path radiance to viewing angle decreases for larger sun zenith angles, as shown in Figure 1 [7].

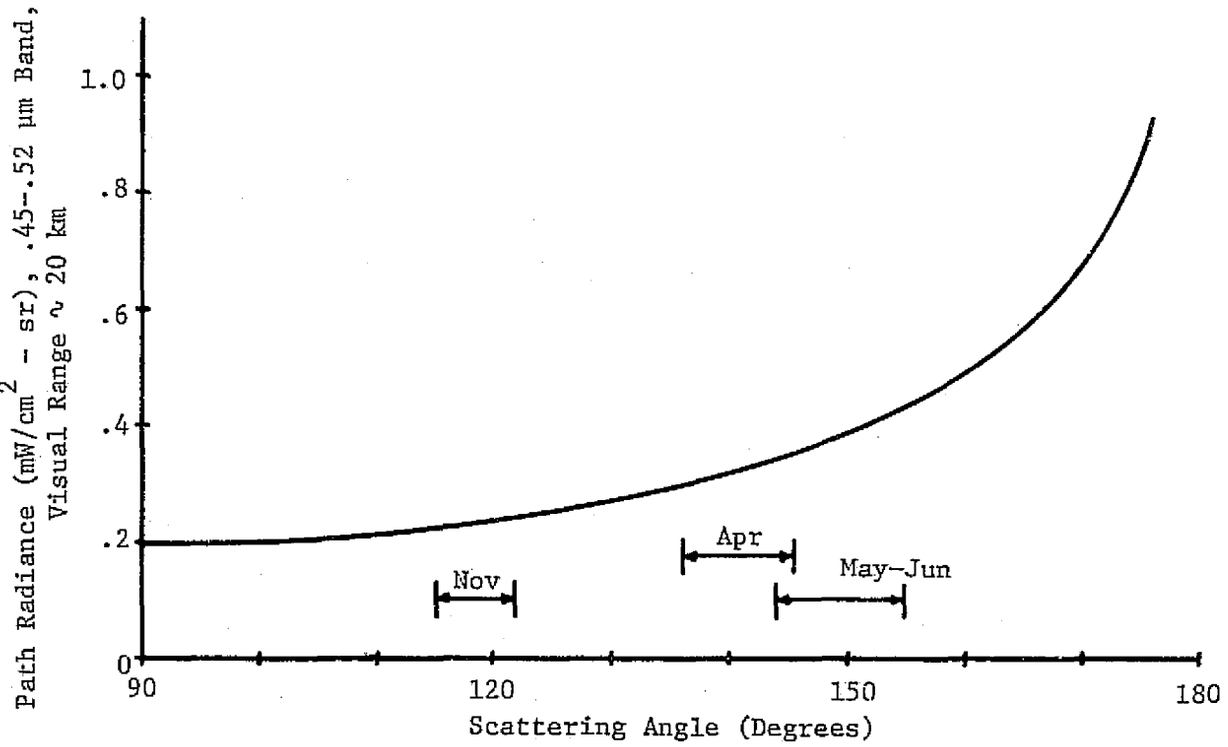


FIGURE 1. GENERAL TREND OF PATH RADIANCE AS A FUNCTION OF SCATTERING ANGLE. Scattering-angle differences for a simulated $\pm 6^{\circ}$ change in view angle are indicated. Vertical scale and detailed curve slope depend on atmospheric condition and spectral band specifications.

In order to simplify manipulations of Equation 2, the equation may be written as follows

$$L = A\left[\frac{4\mu_o^2\rho}{\Delta}\left(K_o\frac{\rho_r}{\rho} + K_1\right) + K_2\right] \tag{10}$$

with

$$K_o = C_o e^{-\tau/\mu} \tag{11}$$

$$K_1 = \frac{1}{2}\frac{\tau}{\mu}e^{-\tau/\mu} + C_1\left\{1 - \left[1 + \frac{\tau}{\mu}\right]e^{-\tau/\mu}\right\} \tag{12}$$

$$K_2 = C_2\left\{\left[1 - \frac{\tau}{\mu}\right] - e^{-\tau/\mu}\right\} + \frac{\tau}{\mu} \tag{13}$$

Equation 10 still shows explicitly most of the dependence of the radiance, L, on the background albedo, ρ, and the cosine of the sun zenith angle, μ_o, however A and C_2 also are strong functions of μ_o, as shown in Equations 3 and 7.

The signal, x, recorded by a multispectral scanner, given an input signal, L, can be represented by

$$x = GL + \delta \tag{14}$$

with G representing the gain of the scanner and δ representing an additive signal offset. (Scanner noise may be considered to be a time dependent perturbation in G and δ, however the following discussion will assume that scanner noise is small enough to be safely ignored.) Equation 14 represents a linear scanner response, however the same functional form can be used to approximate portions of a nonlinear response, producing a piecewise linear representation. The

quantities x, G, L, and δ are all functions of wavelength or channel number.

Denoting parameters corresponding to a standardized atmospheric state and standardized measurement conditions with primes, and incorporating the form of Equation 10 into Equation 14, one may write two equations to describe the recorded signal levels from a scene before and after standardization, respectively:

$$x = G\Lambda\left[\frac{4\mu_o^2\rho}{\Delta}\left(K_o\frac{\rho_t}{\rho}+K_1\right)+K_2\right]+\delta \tag{15}$$

$$x' = G\Lambda'\left[\frac{4\mu_o'^2\rho'}{\Delta'}\left(K_o'\frac{\rho_t'}{\rho'}+K_1'\right)+K_2'\right]+\delta \tag{16}$$

In Equations 15 and 16 scanner gain, G, and offset, δ, have been assumed to be stable. These two equations may then be combined to obtain a relation between the original signal, x, and the standardized signal, x', by first defining a quantity Q such that

$$Q \equiv \frac{\rho_t'}{\rho_t} \simeq \frac{\rho'}{\rho} \tag{17}$$

The quantity Q is intended to represent the effect of bidirectional reflectance, and is expected to be primarily a function of view angle. Equating ρ_t'/ρ' to ρ_t/ρ as indicated by Equation 17, solving Equation 15 for ρ_t/ρ, and substituting for ρ_t'/ρ' in Equation 16, we obtain

$$x' = Q\frac{\Lambda'\mu_o'^2\Delta K_o'}{\Lambda\mu_o^2\Delta' K_o}x+\left(1-Q\frac{\Lambda'\mu_o'^2\Delta K_o'}{\Lambda\mu_o^2\Delta' K_o}\right)\delta$$
$$+\Lambda'\left(K_2'-Q\frac{\mu_o'^2\Delta K_o'}{\mu_o^2\Delta' K_o}K_2\right)G+\Lambda'\frac{4\mu_o'^2\rho'}{\Delta'}\left(K_1'-\frac{K_o'}{K_o}K_1\right)G \tag{18}$$

Equation 18 is the basic starting point for the development of both the XSTAR and XBAR signature extension preprocessing algorithms. The development of the theory from this point which is pertinent to each of these algorithms is discussed in Sections 5 and 6. Another basic theoretical development pertinent to both these algorithms is the evolution of a haze diagnostic procedure for estimating the optical thickness for an arbitrary Landsat agricultural scene. This is discussed below.

3.3 DEVELOPING A HAZE DIAGNOSTIC

The parameters describing the illumination and viewing geometry for a specified data acquisition can be easily calculated. However, in order to devise a preprocessing technique to standardize physical effects one needs to estimate the remaining factors in the radiative transfer equation, as they appear in Equation 18. The major unknown factors are optical thickness, background albedo, scanner calibration (G and δ, which for a satellite usually change after launch), and atmospheric absorption. Optical thickness is the most significant of these factors. The other factors will be discussed in Sections 5 and 6.

In principal the optical thickness is a separate unknown quantity in each spectral band of a scanner. However, to determine the optical thickness for a single spectral band by analyzing the appearance of the data only within that band usually produces a rather inaccurate result unless the band is one in which all other useful information is essentially nonexistent, or one for which some special scene characteristics are known. By treating the optical thickness as an independent unknown quantity in each spectral band, one in effect is faced with too many unknown quantities. This problem can be rendered more tractable by obtaining a relationship among the optical thicknesses for the various bands. One possible relationship can be determined by assuming that the optical thickness in each spectral band is a

linear function of the amount of haze in the atmosphere. If we denote the amount of haze in the standardized condition by γ', and the amount of haze in an observed condition by $\gamma' + \gamma$, then for a scene before and after standardization, respectively, we may write

$$\tau = \tau_R + \alpha(\gamma' + \gamma) \tag{19}$$

$$\tau' = \tau'_R + \alpha\gamma' \tag{20}$$

The parameters γ' and γ are scalar quantities (independent of wavelength) characterizing the amount of haze in the atmosphere. In effect γ' and γ are measures of the aerosol optical thickness at some standard wavelength, for which $\alpha = 1$. The parameter α is a function of wavelength, having a unique value for each spectral band. For Landsat data we have defined

$$\alpha \equiv \begin{pmatrix} 1.2680 \\ 1.0445 \\ .9142 \\ .7734 \end{pmatrix} \tag{21}$$

The values chosen for α are based on the relative magnitude of the aerosol optical thickness in each of the Landsat bands 4 through 7, for an atmosphere with a horizontal visual range of 23 km (a relatively clear atmosphere). Remembering that $\tau_R = \tau'_R$ (Rayleigh optical thickness is independent of atmospheric condition), we may write

$$\tau = \tau' + \alpha\gamma \tag{22}$$

Using the relation defined by Equation 22 in Equation 18, and after specifying the other factors of Equation 18, we obtain a definition of multiplicative and additive changes to Landsat signals as a

function of a single parameter, γ, which is to be determined. After finding some measurable change in Landsat signal values which is a monotonic function of γ, our haze diagnostic can then be specified, since we need only determine the value of γ which will change the observed Landsat signals to correspond to the signal configuration characterizing the standardized condition. This step has been accomplished, using our knowledge of the principal components of Landsat agricultural data distributions, as described below.

It has been noted that Landsat agricultural data tends to occupy a region of the signal space which has a form similar to a Tasselled Cap [8]. This distribution is flattened so that the first two principal components of the data distribution define a hyperplane containing most of the variance of the data. In the Tasselled Cap model, two specially oriented axes lying within this hyperplane have been labeled the soil brightness axis and the green development axis. The third most significant principal axis has been labeled "yellow stuff", while the fourth has been called "non-such". The labels for the axes are based on the features of the data which appear to be most highly correlated with each axis. The axial directions characterizing the Tasselled Cap description of Landsat I data have been found to differ by up to five degrees from what would appear to be equivalent axial directions for Landsat II data. Since the great majority of the data available for our analysis was Landsat II data, a more highly tuned Tasselled Cap description for Landsat II was needed before atmospheric effects on the Tasselled Cap distribution could be readily determined.

The determination of the Tasselled Cap axes appropriate for Landsat II data was accomplished in two steps. The first step attempted to define a pair of mutually orthogonal two dimensional subspaces such that the first (or major) hyperplane contained as much of the variance of the Landsat II data as possible, while the second (or minor) hyperplane contained as little of the variance as possible. For this analysis

signatures were calculated from the data distributions of 10 Landsat II data sets, comprising 5 LACIE segments in Oklahoma, 3 in Kansas, 1 in Texas, and 1 in Arizona, all recorded during the month of April 1975. Since our goal, for the present, was to devise a standardization to an average atmospheric condition (rather than a perfectly clear atmospheric condition), an average orientation of the major and minor hyperplanes was sought. Hence, a procedure was devised for determining the average orientation of the major and minor hyperplanes from the eigenvectors of the 10 data sets. This procedure began with an initial estimate for the Tasselled Cap axes, designated as column vectors within the rotation matrix R_I (denoted by R in Reference 8), which has been used for the Landsat I fixed linear Tasselled Cap transformation. Each of the 10 sets of 4 eigenvectors, designated as column vectors within a rotation matrix R_E, could then be compared with R_I. To do this, first a matrix T was calculated such that

$$T \equiv R_E^T R_I \tag{23}$$

Since the orientation of the first two eigenvectors within the major hyperplane is quite variable, while the third and fourth eigenvectors are usually consistently oriented, the expected form for T would be represented by

$$T \simeq \begin{pmatrix} \pm \cos\theta & \pm \sin\theta & 0 & 0 \\ -\sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \equiv \hat{T} \tag{24}$$

Since the polarity of an eigenvector is an artifact of the routine which calculates it, either the plus or the minus sign could be appropriate for the first row of the $\hat{T}$ matrix (Equation 24), while the minus sign in the third column and row of the matrix was found always to be appropriate. The $\hat{T}$ matrix is a rotation matrix which retains the orientation of the major and minor hyperplanes of R_E while changing the orientation of the axes within the major hyperplane to correlate optimally with the axes of R_I. Thus, a Tasselled Cap rotation matrix for Landsat II data ($\hat{R}_{II}$) could be estimated by replacing T in Equation 23 with $\hat{T}$ and then multiplying both sides of the equation from the left by R_E, producing

$$\hat{R}_{II} \equiv R_E\hat{T} \approx R_I \tag{25}$$

The correlation of the axes of $\hat{R}_{II}$ with the axes of R_I is obtained in the least squares sense when θ is defined such that

(1) if $T_{11}T_{22} \geq T_{12}T_{21}$:

the plus signs in the first row of $\hat{T}$ are used, and

$$\theta = \arg(T_{12} - T_{21}) \left[\frac{\pi}{2} - \arg(T_{11} + T_{22}) \tan^{-1} \left| \frac{T_{11} + T_{22}}{T_{12} - T_{21}} \right| \right] \tag{26}$$

(2) if $T_{11}T_{22} < T_{12}T_{21}$:

the minus signs in the first row of $\hat{T}$ are used, and

$$\theta = -\arg(T_{12} + T_{21}) \left[\frac{\pi}{2} + \arg(T_{11} - T_{22}) \tan^{-1} \left| \frac{T_{11} + T_{22}}{T_{12} - T_{21}} \right| \right] \tag{27}$$

In this manner, estimates for $\hat{R}_{II}$ were obtained from each eigenvector matrix, R_E. Inspection of these results revealed that the eigenvector matrices for two of the ten data sets indicated "yellow stuff" and "non-such" axes which differed significantly from the trend in the remaining eight data sets. One of these two data sets (Segment 1239, Noble Co., Oklahoma) had so little variance along the "yellow stuff" axis that the eigenvector was ambiguous, while the other of the two data sets (Segment 1316, Yuma Co., Arizona) had abnormally high variance along the "non-such" axis, which appeared to be a rare instance of useful information correlating with the "non-such" direction. (The physical meaning of this "non-such" axis has not yet been determined.) These two estimates for $\hat{R}_{II}$ were then set aside, and the remaining eight estimates for $\hat{R}_{II}$ were averaged, component by component. This resulted in an average estimate for R_{II} whose components were no longer orthonormal. This average estimate for R_{II} was then orthonormalized using the standard Gram-Schmidt procedure, beginning with the soil brightness vector and then proceeding to the green development vector, followed by the "yellow stuff" and "non-such" vectors.

The second step in determining the Tasselled Cap matrix, R_{II}, for Landsat II data was to perform separate rotations within the major and minor hyperplanes to optimize the within-plane orientation of the Tasselled Cap axes. Angles θ_1 ($-5^o42'$) and θ_2 ($0^o46'$) were defined for the two rotations such that the "green" and "non-such" components of the special signal vector x^* (defined in Section 5) would be zero. Letting R'_{II} denote the orthonormalized average of the $\hat{R}_{II}$ matrices, a rotation matrix T'

$$T' \equiv \begin{pmatrix} \cos\theta_1 & \sin\theta_1 & 0 & 0 \\ -\sin\theta_1 & \cos\theta_1 & 0 & 0 \\ 0 & 0 & \cos\theta_2 & \sin\theta_2 \\ 0 & 0 & -\sin\theta_2 & \cos\theta_2 \end{pmatrix} \qquad (28)$$

was defined which was used to calculate R_{II} according to Equation 29.

$$R_{II} = R'_{II} T' \tag{29}$$

The resulting Landsat II Tasselled Cap transformation matrix is given in Equation 30.

$$R_{II} = \begin{pmatrix} .33231 & -.28317 & -.89952 & -.01594 \\ .60316 & -.66006 & .42830 & .13068 \\ .67581 & .57735 & .07592 & -.45187 \\ .26278 & .38833 & -.04080 & .88232 \end{pmatrix} \tag{30}$$

This orientation of the Landsat II Tasselled Cap axes has been found to be particularly suitable for determining and applying a haze diagnostic procedure (as was intended). However, an analysis of the variability in Landsat II signals from bare soils in Kansas has determined that the first principal component of this soil variability, which contains approximately 95% of the total observed variance, is within 1 degree of alignment with the "brightness" direction defined by Equation 30, after this soil principal component is projected onto the "brightness-greenness" hyperplane. (This projection would remove any rotation of the bare soil principal component out of the "brightness-greenness" hyperplane which could have been caused by atmospheric haze.) Hence, the correlation of the "brightness" and "green" directions with soil brightness and green development appears to have been retained in the R_{II} matrix (Equation 30).

Displaying Landsat II Tasselled Cap transformed data distributions in the coordinates "brightness" vs. "yellow", we have observed that while the scatter of the data out of the hyperplane in the "yellow" direction is usually very small, the hyperplane shifts and rotates in a clearly discernible manner which is correlated with the atmospheric

condition. This motion of the data hyperplane in "brightness" vs. "yellow" provides the basis for a haze diagnostic procedure. Specific details of the procedure applicable to the XSTAR and XBAR algorithms are presented in Sections 5 and 6.

4

AUTOMATIC SCREENING OF LANDSAT MSS DATA

Not all Landsat data lies within a well defined hyperplane as described in Section 3.3. In particular, garbled data or data from clouds, localized dense haze concentrations (diffuse clouds), cloud shadows, snow, or water can appear to be atypical and can lead to errors in calculating the haze diagnostic(s) required by a preprocessing algorithm such as XSTAR or XBAR. Hence, it was necessary to develop a data screening procedure to edit out confusing data before XSTAR or XBAR could be most effectively applied. For this application of data screening, errors of commission in identifying confusing data are acceptable, provided that enough data remains to characterize the atmospheric condition with sufficient accuracy.

Data screening can also be used to edit the input to a classifier. For this purpose errors of commission and errors of omission from the screening process both need to be minimized. Since there is a temptation, if not an outright desire, to use data screening both for obtaining a better haze diagnostic and for editing the input to a classifier, regardless of any initial limited intent for the screening process, an attempt was made to develop a screening procedure (called SCREEN) that would adequately suit both of these needs and yet could be applied with minimum supervision.

The SCREEN algorithm uses thresholds on linear combinations of Landsat data values, after applying a cosine sun zenith angle correction, to separate regions of the data space which are of interest. To determine the most appropriate screening thresholds, we needed to develop an understanding of the physical interpretation of data within typical Landsat data distributions. This insight was gained through experience with the Tasselled Cap data transformation [8]. Hence, the first step of the screening procedure is to transform the signal

vector, x, for each Landsat II pixel to obtain the corresponding sun angle corrected Tasselled Cap vector, z:

$$z = \frac{\mu_o'}{\mu_o} R_{II}^T x \tag{31}$$

with μ_o and μ_o' representing the cosine of the sun zenith angle for the data acquisition and for the standardized condition, respectively. In this case we have chosen

$$\mu_o' \equiv \cos 39^o \tag{32}$$

which is typical for Landsat data acquired in April in Kansas. The Landsat II Tasselled Cap rotation matrix, R_{II}, is discussed in Section 3.3 and is defined in Equation 30.

The next step of the SCREEN procedure is to circumscribe the usual Landsat data distribution, using several separate linear thresholds, and to label any pixels with outlying signal vectors as garbled data. The remaining "good" data is then split up into separate, mutually exclusive subregions to identify in succession dense clouds (or snow), diffuse clouds (or localized dense haze concentrations atypical of normal Landsat scenes), water, and cloud shadows. The location of these screening thresholds has been determined by studying 13 LACIE acquisitions from North Dakota and Montana and 19 LACIE acquisitions from Kansas, carefully selected to be examples of particular screening problems. A condensed, detailed programmer's description of the resulting SCREEN algorithm is presented in Reference 9.

A few of the SCREEN thresholds are shown in Figures 2 and 3. These figures display the screening thresholds in the Tasselled Cap rotated data space, without any offset applied to the origin (as defined by Equation 31). The regions outside the enclosed areas in

ORIGINAL PAGE IS
OF POOR QUALITY

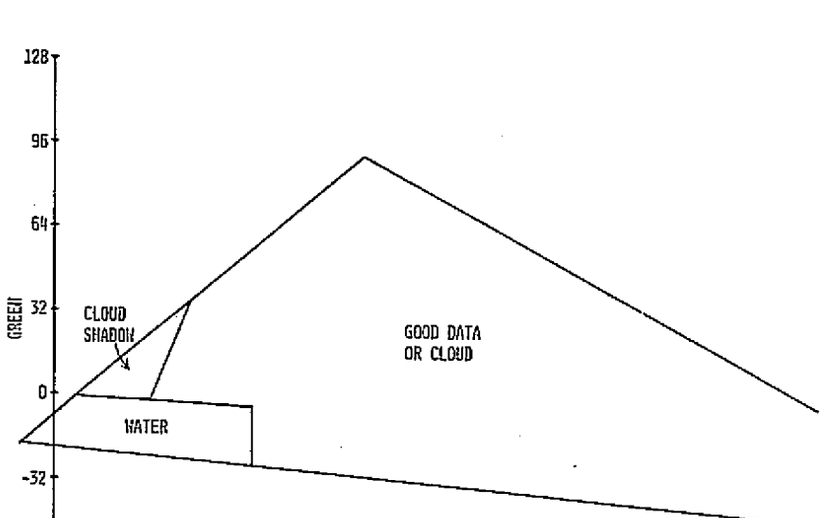


FIGURE 2. SCREEN THRESHOLDS IN TASSELLED CAP ROTATED LANDSAT II DATA SPACE. (Data standardized to 39° sun zenith angle.)

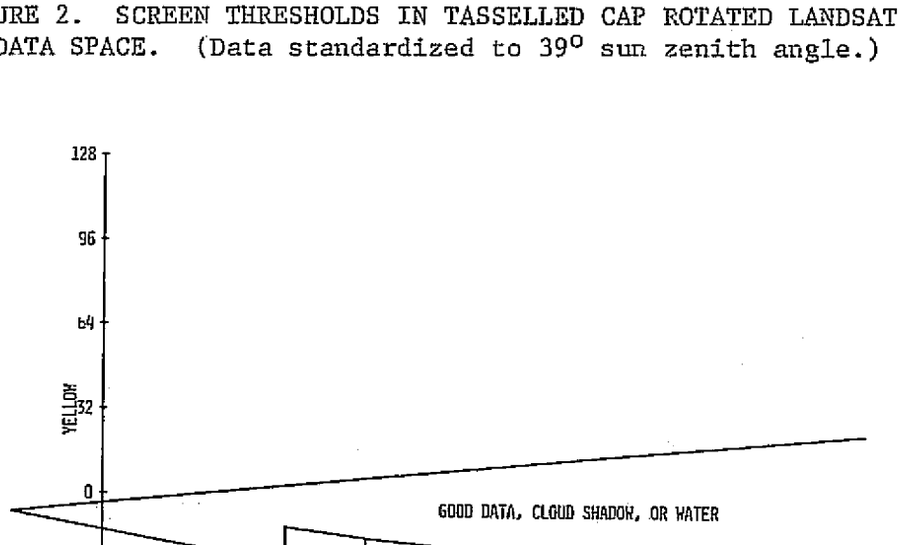


FIGURE 3. SCREEN THRESHOLDS IN TASSELLED CAP ROTATED LANDSAT II DATA SPACE. (Data standardized to 39° sun zenith angle.)

the figures correspond to garbled data values. The SCREEN thresholds were found to be rather sensitive. Striping effects in the Landsat data were often sufficient, especially for low sun elevations, to cause data values to cross the threshold boundaries. Better Landsat radiometric consistency, smaller digitization intervals (more significant bits), and additional data channels (e.g., thermal data) all would help to make this type of data screening procedure more effective.

Figure 4 shows a classification map generated from the output of the SCREEN algorithm. The symbols in the map are assigned as follows: B for garbled ("bad") data, C for dense clouds, H for diffuse clouds (dense haze), W for water, S for cloud shadows, and N for cloud shadow over water. Note the clear definition of the river which runs from top to bottom through the scene and the definition of the cloud and cloud shadow areas. Cloud areas are to the right and slightly below the corresponding shadow areas in the figure. The clouds are at various altitudes, hence the displacement of clouds from their shadows varies throughout the scene. Some areas of confusion between cloud shadow and water are present in this scene. For instance, the areas classified as cloud shadow near line 20, pixel 90, and near line 40, pixel 116, are actually lakes. Similar misclassifications observed in another scene for which ground truth was available indicated that such lakes are actually shallow water with vegetation (e.g., grass) growing up through the water. Such areas are indistinguishable from cloud shadows, using Landsat spectral data alone. Near line 95, pixel 53, is an area of cloud shadow misclassified as water. Within the extensive cloud shadow area around line 50, pixel 150, are several pixels identified as cloud shadow over water. Note the every-sixth-line structure of these latter misclassification areas. Better destriping of the Landsat data would be a partial remedy for such problems.

ORIGINAL PAGE IS
OF POOR QUALITY

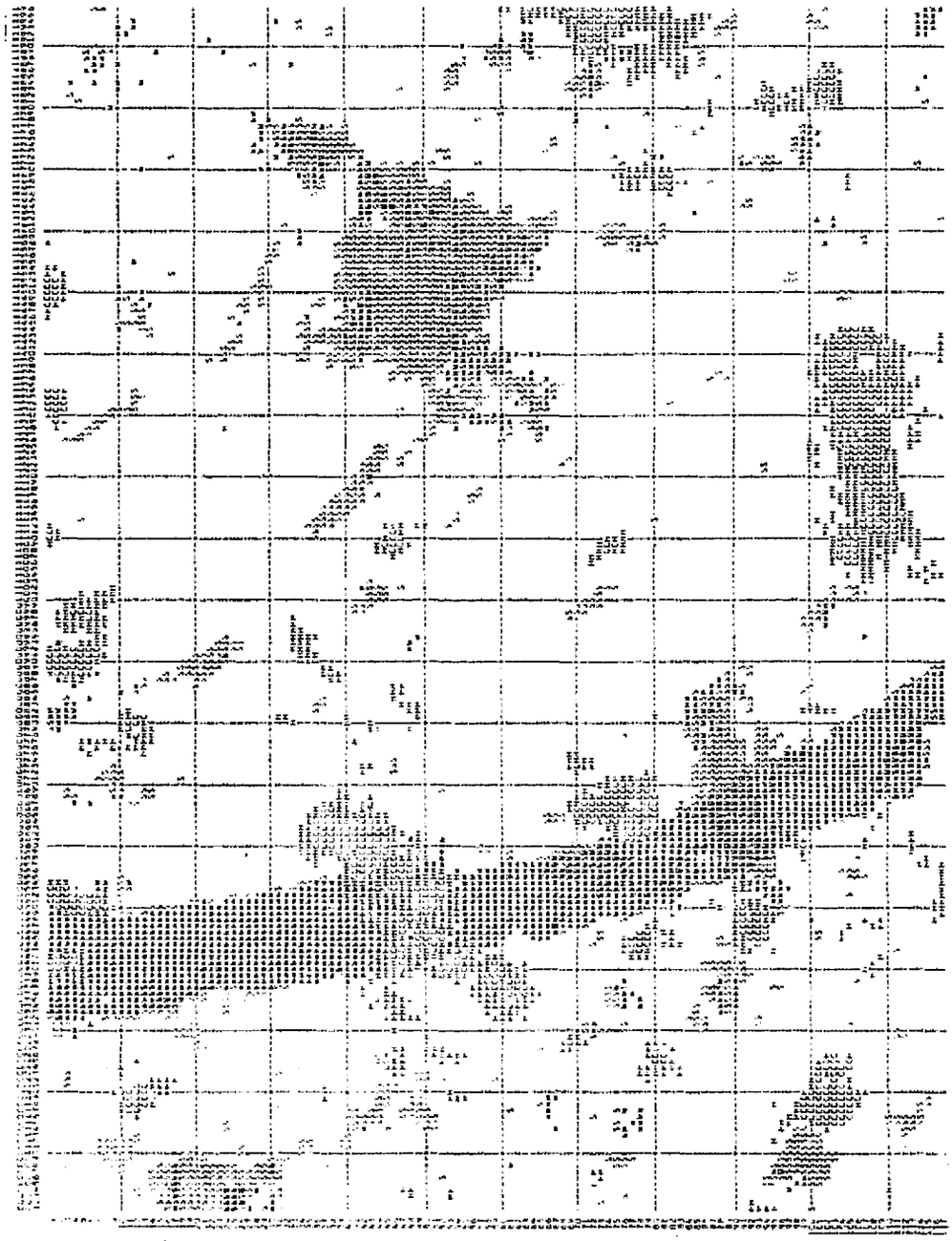


FIGURE 4. SCREEN CLASSIFICATION MAP. LACIE Segment No. 1610, Bottineau Co., N.D., 21 June 1975 (75172).

Figure 5 shows a screening classification map for a scene which is so overcast that it ordinarily would not be processed. However, this scene serves as a good example for how garbled (or "bad") data can be detected. Repeated scan lines, fill data, and bit slips are three of the normal causes for "bad" data in Landsat images, but in this scene two other sources of garbled data were detected and noted. The first cause was due to the band 4 value changing to zero for one or two isolated pixels, while the signals in the other bands (5-7) remained similar to the surrounding signal values. The second cause was due to the band 4 signal increasing by approximately 20 counts (∿40%), while the signals in the other bands again remained at typical values. Both problems appear to have come from a single band 4 detector or from its associated circuitry or ground processing. Neither of these problems would have been easily spotted by examining usual film products for this scene, however the SCREEN algorithm was able to identify these problems routinely (and automatically).

The ERIM SCREEN procedure is somewhat more refined than a simpler procedure recently developed by the Agricultural Research Service [10] which uses only bands 5 and 7. Dense coverings of snow are classified by the ERIM procedure as dense cloud, while separation of water from other categories is about as accurate as the spectral data by itself will permit. There is some tendency for false alarms to increase at lower sun elevations, mostly due to the effects of striping in the Landsat data. Some improvement in the separation of clouds from bright fields could be obtained by allowing a few of the screening thresholds to be adjusted separately for each scene by user interaction. However, as it stands, the SCREEN procedure is reasonably effective both for removing confusing data from haze diagnostic calculations and for editing the input to a classifier, without supervision. We do recommend, nevertheless, that users monitor its performance visually.

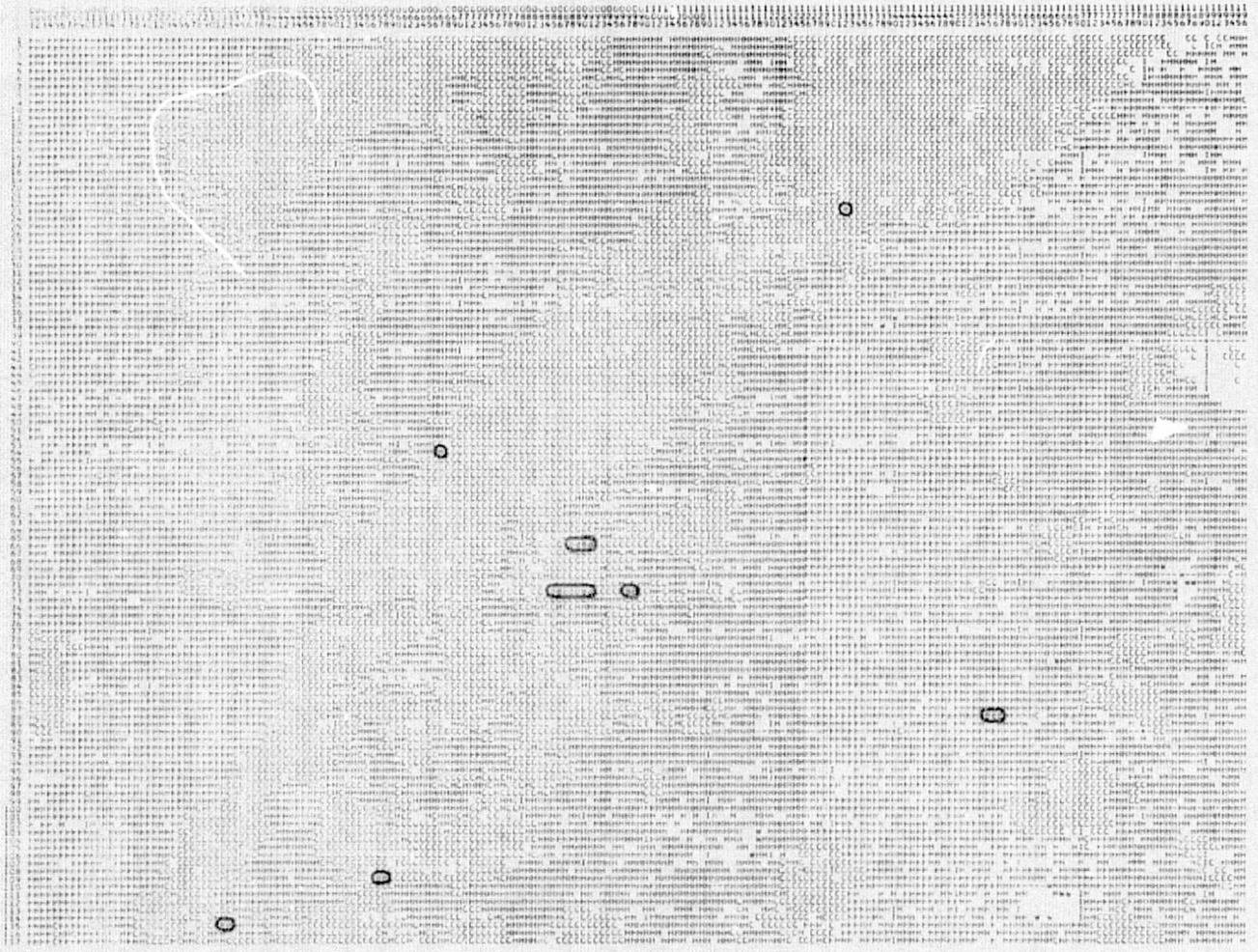


FIGURE 5. SCREEN CLASSIFICATION MAP. LACIE Segment No. 1553, Carter Co., Montana, 15 August 1975 (75227). Garbled data is circled.

ERIM
FORMERLY WILLOW RUN LABORATORIES, THE UNIVERSITY OF MICHIGAN
ORIGINAL PAGE IS OF POOR QUALITY

5

THE XSTAR SIGNATURE EXTENSION PREPROCESSING ALGORITHM

5.1 DERIVATION

The XSTAR signature extension preprocessing algorithm is the result of a mixture of physical intuition, empirical observation, and a greatly simplified formulation based on the ERIM radiative transfer model. Mathematically, the XSTAR algorithm may be derived as follows.

In Section 3.2 we noted that the quantities Λ, Δ, C_o, C_1, and C_2 in the radiative transfer model (Equation 2) usually varied by no more than $\pm 5\%$ with changes in τ (optical thickness) and ρ (background albedo). We may also note that the cosine of the viewing angle relative to nadir (μ or μ') cannot vary by more than 0.6%, due to the limited scan angle of the Landsat satellite ($\pm 6^o$). Hence, using Equation 22, and referring to Equation 11, we may write

$$\frac{K'_o}{K_o} \simeq e^{\alpha\gamma} \tag{33}$$

If we further assume that $\mu_o \simeq \mu'_o$ (that changes in sun zenith angle are small) and that $Q \simeq 1$ (that variations in bidirectional reflectance are small), Equation 11, relating the standardized signal x' to the observed signal x, may be simplified to the following form

$$x' \simeq e^{\alpha\gamma}x + (1 - e^{\alpha\gamma})\delta + \Lambda'(K'_2 - e^{\alpha\gamma}K_2)G + \Lambda' \frac{4\mu'^2_o\rho'}{\Delta'}(K'_1 - e^{\alpha\gamma}K_1)G \tag{34}$$

The terms of Equation 34 which involve the scanner gain, G, can be expanded into a power series in ascending powers of $\alpha\gamma$. Since to the level of approximation assumed in this derivation this power series

has no term involving $(\alpha\gamma)^0$, Equation 34 may be rewritten as

$$x' \simeq e^{\alpha\gamma}x + (1 - e^{\alpha\gamma})\delta + a_1(\alpha\gamma)^1 + a_2(\alpha\gamma)^2 + \ldots \qquad (35)$$

Approximating $-(\alpha\gamma)^1$ by $(1 - e^{\alpha\gamma})$, and modifying the polynomial coefficients a_2, a_3, etc., accordingly, Equation 35 may then be restated as

$$x' \simeq e^{\alpha\gamma}x + (1 - e^{\alpha\gamma})(\delta - a_1) + P(\alpha\gamma) \qquad (36)$$

The quantity a_1 is a function of the scanner gain, G, and of all the radiative transfer equation variables (μ_o', ρ', τ', etc.) characterizing the standardized condition. The polynomial function $P(\alpha\gamma)$ is a function of these same variables, with its first term proportional to $(\alpha\gamma)^2$, and thus represents higher order effects of changes in optical thickness.

The XSTAR algorithm is based on the mathematical form of Equation 36, excluding the higher order terms represented by $P(\alpha\gamma)$. To define the algorithm one needs to estimate the value of $(\delta - a_1)$ for each spectral band. This has been done empirically for a restricted data set, as described below.

First note that Equation 36 describes a multiplicative and additive change applied to a single channel signal value, x, to obtain the corresponding standardized signal, x'. The standard form of this transformation is

$$x' = A\,x + B \qquad (37)$$

with the scalar quantities A and B in Equation 37 representing multiplicative and additive factors, respectively.

For any such multiplicative and additive transformation with a multiplicative factor not equal to unity, one can define a unique

signal value x^*, in each channel by

$$x^* \equiv \frac{B}{1 - A} \tag{38}$$

where A, B, and x^* are all functions of wavelength or channel number. Note that the signal value x^* is invariant under the given transformation. The standard transformation (Equation 37) may be rewritten as

$$x' = A\,x + (1 - A)x^* \tag{39}$$

or as

$$x' - x^* = A(x - x^*) \tag{40}$$

Equation 40 indicates that the value of x^* in each data channel specifies a point or origin in the signal space relative to which the remainder of the signal space expands or contracts according to the effect of each multiplicative factor. Comparing Equation 36, excluding $P(\alpha\gamma)$, to Equation 39, it is apparent that x^* can be equated to $(\delta - a_1)$ in Equation 36. The form of the physical effects standardization then becomes

$$x' \simeq e^{\alpha\gamma}x + (1 - e^{\alpha\gamma})x^* \tag{41}$$

The existence of the special signal value, x^*, has led to the name XSTAR for the resulting preprocessing algorithm. (Note that the sign convention chosen for γ is opposite to that chosen in previous documentation of the XSTAR algorithm [11]. This sign convention has been changed in order to be consistent with the XBAR presentation in Section 6 and to define γ as a scalar parameter monotonically related to the amount of haze present in the scene to be preprocessed.)

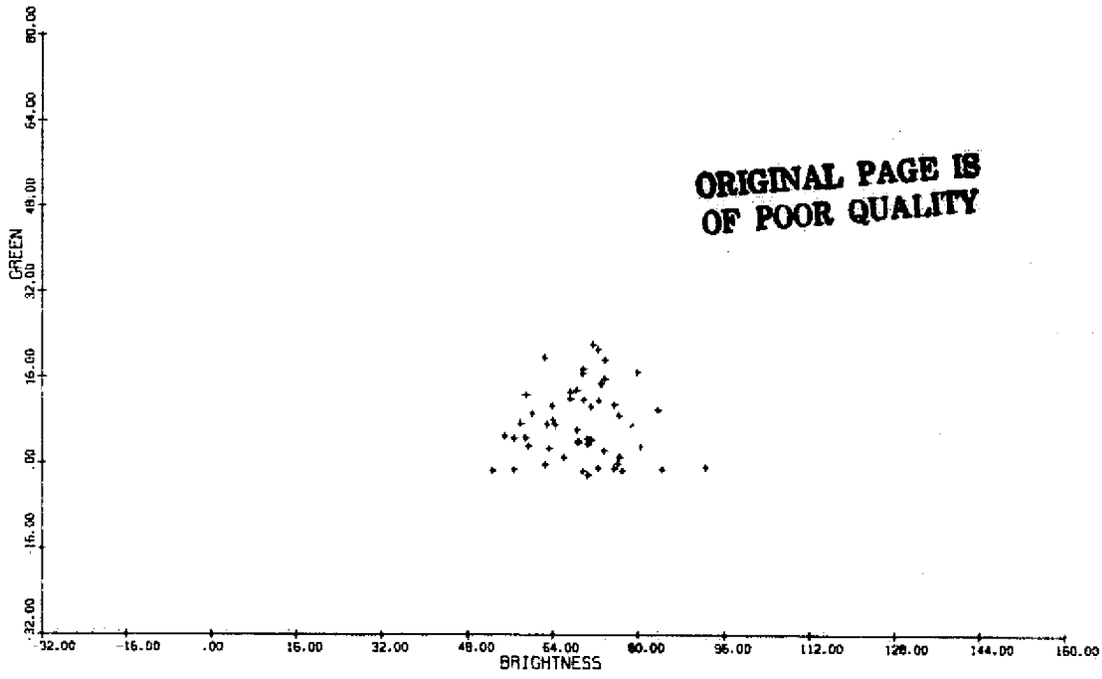
Two fortuitous circumstances with respect to Landsat data have made the task of developing preprocessing algorithms to standardize

physical effects less difficult than it might have been. The first is the occurrence of areas of overlap in the ground swath covered by Landsat on consecutive days. The second is the occurrence of "redundant" information in the Landsat bands 4 through 7. This "redundant" information causes the Landsat data to lie in a hyperplane and is what has made a reliable haze diagnostic procedure possible. On the other hand, consecutive day Landsat acquisitions for selected scenes have provided the controlled observation conditions necessary for studying physical effects on Landsat signals.

The value of x^* in each Landsat band has been estimated by optimizing the performance of the XSTAR algorithm on 10 consecutive day data sets. These data sets are the same ones that were used to determine the Landsat II Tasselled Cap rotation matrix, R_{II}, and are described in Section 3.3. All 20 of these acquisitions had solar zenith angles of $40^{\circ} \pm 4^{\circ}$. At first γ was allowed to assume whatever value was necessary to match the data from one day of each data set to the other day. After a stable estimate for x^* was obtained (partly by trial and error), the final formulation of a haze diagnostic procedure was possible.

5.2 THE XSTAR HAZE DIAGNOSTIC

As is commonly known, the effect of increasing haze on MSS signals is to reduce the available signal contrast (dynamic range encompassed by the data) and to offset most signals toward brighter levels. Objects which are especially bright, however, may appear somewhat dimmer after a haze increase. This qualitative observation is illustrated by Figures 6 and 7, which show distributions of cluster means from Landsat II MSS data for two acquistions of a LACIE sample segment obtained on consecutive days. These data distributions are displayed in the Tasselled Cap rotated coordinates "green" vs."brightness". Analyst interpreters, examining false color film images generated from the MSS data for these scenes, described the atmospheric conditions as "clear" on the first day (Figure 6),



ORIGINAL PAGE IS
OF POOR QUALITY

FIGURE 6. CLUSTER MEANS FROM A CLEAR DAY (Tasselled Cap Green vs. Brightness) LACIE Segment No. 1178, Bourbon Co., Kansas, 20 April 1975 (75110)

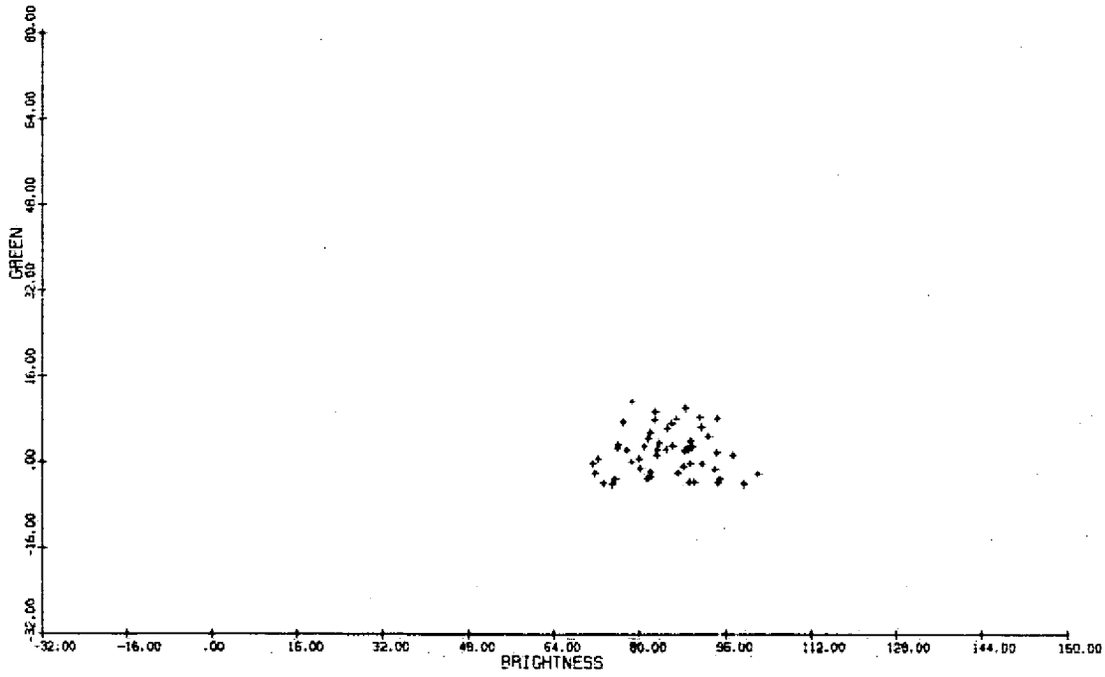


FIGURE 7. CLUSTER MEANS FROM A HAZY DAY (Tasselled Cap Green vs. Brightness) LACIE Segment No. 1178, Bourbon Co., Kansas, 21 April . 75 (75111)

and as "hazy, with some clouds" on the second day. Note that on the second (hazy) day (Figure 7) the data distribution is more compact and "brighter" than on the "clear" day (Figure 6). This same qualitative haze effect is predicted by the XSTAR algorithm, as illustrated schematically in Figure 8. In this figure, the standardized condition is represented by a simulated "soil line" and "green arm", labeled by the letter "B". Condition A then represents a data distribution for a very clear atmospheric condition, while conditions C, D, and E represent progressively hazier conditions. The range of conditions simulated

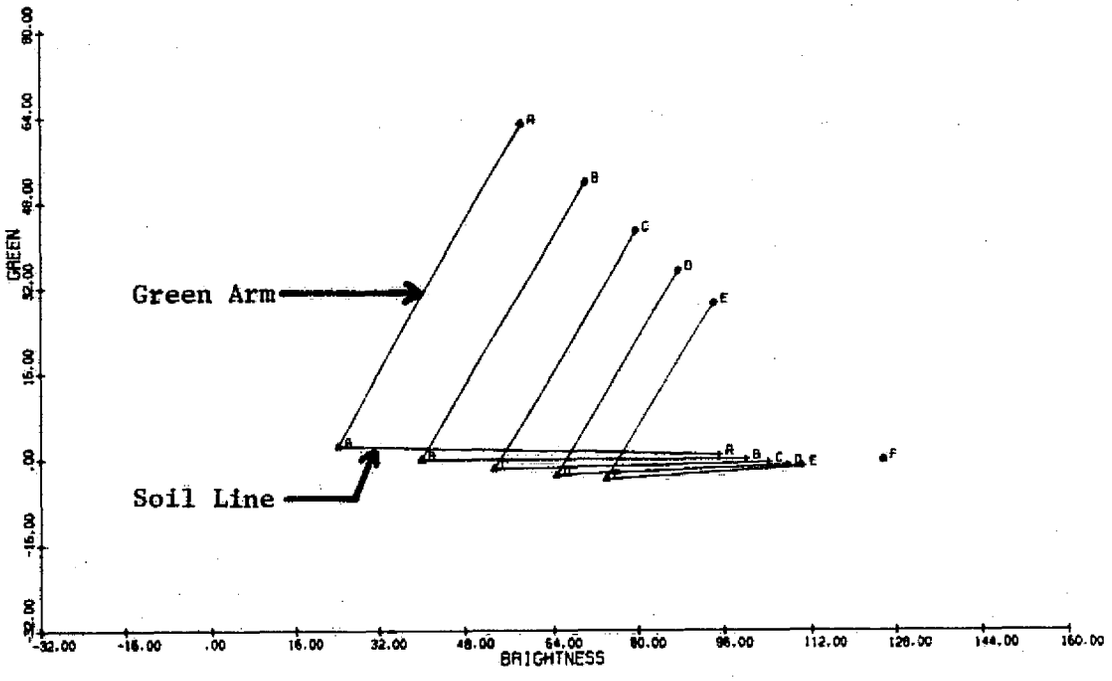


FIGURE 8. EFFECT OF INCREASING HAZE PREDICTED BY XSTAR MODEL
(Tasselled Cap Green vs. Brightness)

is slightly greater than is usually encountered in Landsat data. According to the XSTAR model, increasingly hazy conditions would continue the trend of cases A through E, until the asymptotic limit "F" was reached. This point "F" is the special signal value x^*. In this way one may think of the point x^* as an apparent "point of all haze" to which all data distributions would collapse once the haze became dense enough to reduce the signal contrast to zero. Actually the point x^* is only an apparent "point of all haze", since the effect of the neglected term $P(\alpha\gamma)$ in Equation 36 is to shift the location of the point x^* toward brighter signal values as the haze increases. For the XSTAR algorithm, however, a fixed location for the point x^* has been chosen which produces preprocessing results that are reasonable for normal haze variations.

Figures 9 and 10 display the same distribution of cluster means as Figures 6 and 7, respectively, but in the Tasselled Cap rotated coordinates "yellow" vs. "brightness". Note first of all how little scatter there is of the data about the brightness-greenness hyperplane. This small amount of scatter is typical of Landsat agricultural data distributions. In Figure 10, representing the hazier condition, slightly more scatter is apparent about the plane than in Figure 9, however this scatter is almost entirely due to variations in the haze density within this scene. Note that the data distribution of Figure 10 is offset toward less "yellow" signal values, and is rotated slightly, relative to the data for the "clear" scene in Figure 9. The corresponding motion of the brightness-greenness hyperplane predicted by the XSTAR algorithm is shown schematically in Figure 11. In this case condition B can be observed to correlate well with Figure 9, while condition D, representing an increase in haze (and in γ), correlates well with Figure 10. Note that the slight rotation of the hyperplane in Figure 10 is not predicted by the XSTAR algorithm. This rotation of the hyperplane is not yet fully understood and is not predicted

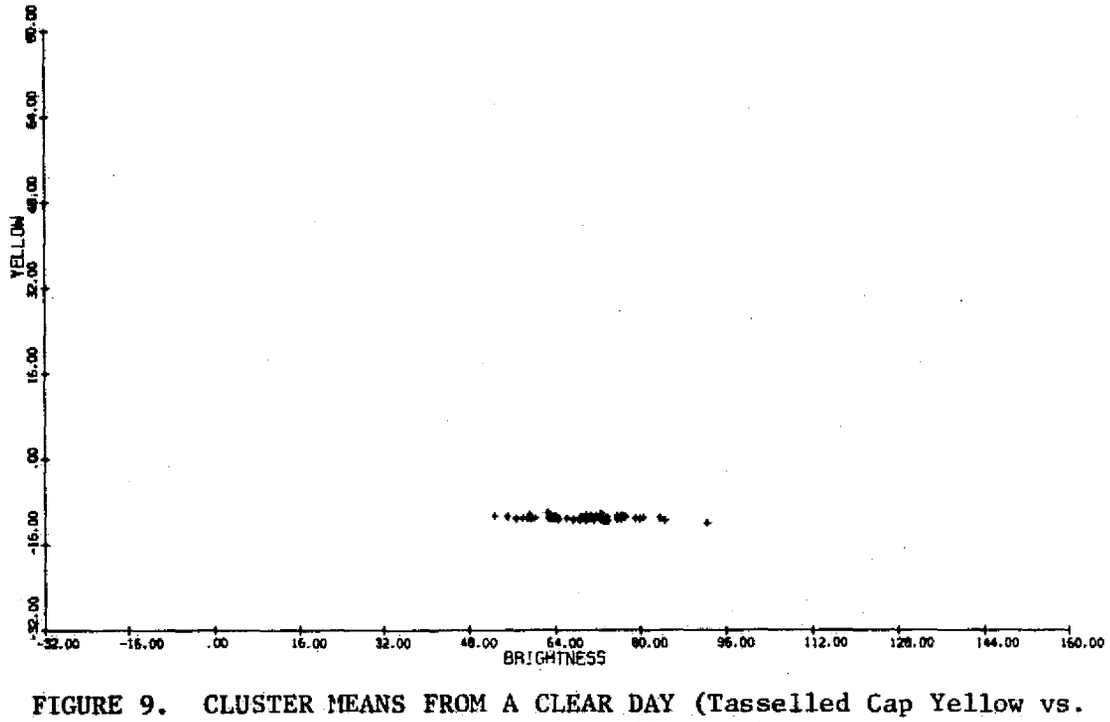


FIGURE 9. CLUSTER MEANS FROM A CLEAR DAY (Tasselled Cap Yellow vs. Brightness) LACIE Segment No. 1178, Bourbon Co., Kansas, 20 April 1975 (75110)

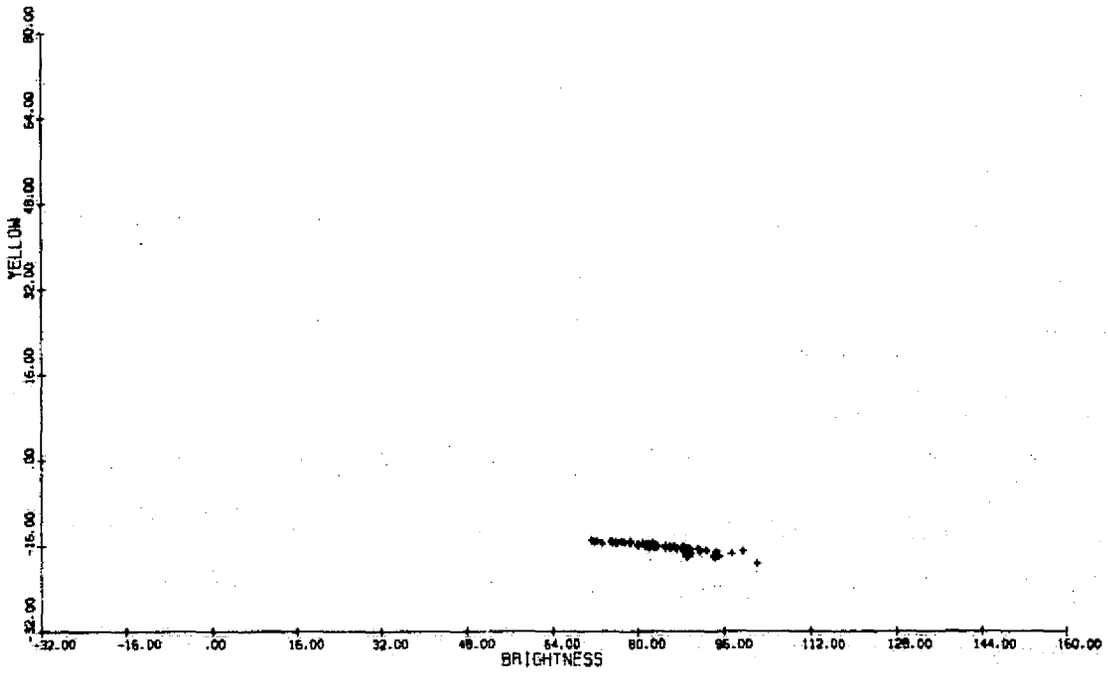


FIGURE 10. CLUSTER MEANS FROM A HAZY DAY (Tasselled Cap Yellow vs. Brightness) LACIE Segment No. 1178, Bourbon Co., Kansas, 21 April 1975 (75111)

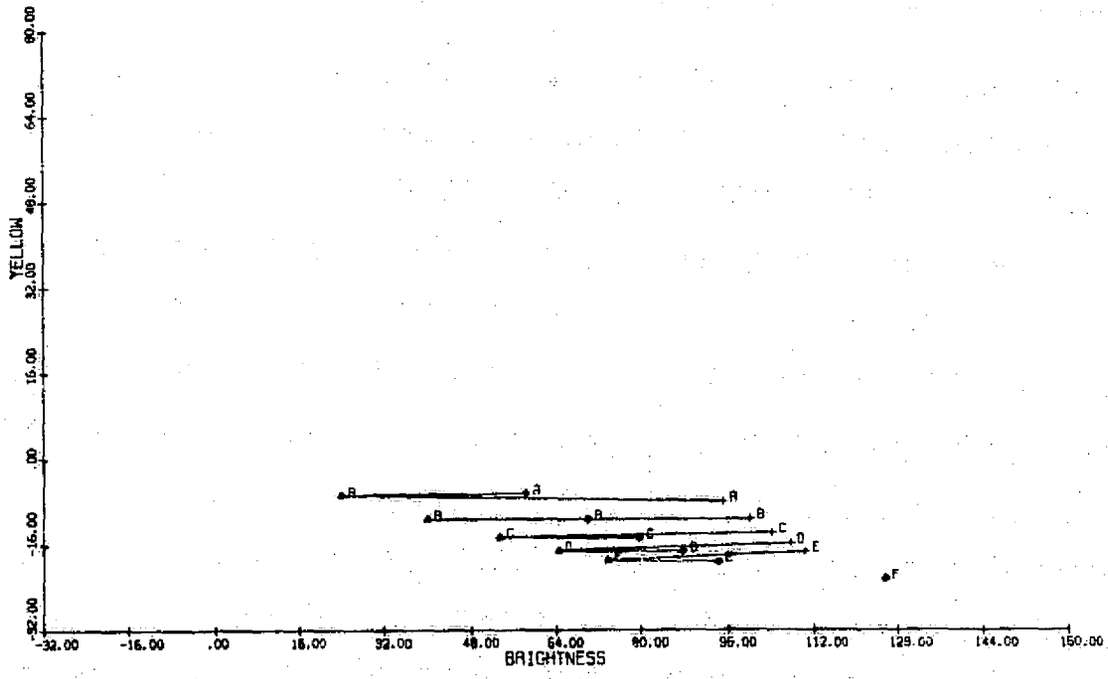


FIGURE 11. EFFECT OF INCREASING HAZE PREDICTED BY XSTAR MODEL (Tasselled Cap Yellow vs. Brightness)

even by more accurate radiative transfer models. We now suspect that this rotation of the Landsat data hyperplane may be related to non-linear scanner performance or to inconsistent calibration of the data (correlated, however, with ambient signal levels), however other possible causes are also conceivable.

The XSTAR haze diagnostic procedure is based on the translational movement of the Landsat brightness-greenness hyperplane in the Tasselled Cap "yellow" direction. Specifically, γ is estimated such that the average "yellow" value for the acquisition to be preprocessed will be transformed to the average "yellow" value characterizing the standardized condition. The standardized "yellow" value (-11.2082 counts) has been chosen to be typical of an average Landsat scene, and is represented

by condition B in Figures 8 and 11. (The data distribution shown for the clear day in Figures 6 and 9 is very close to this standardized condition.) A condensed, detailed programmer's description of the XSTAR signature extension preprocessing algorithm and of its haze diagnostic procedure is presented in Reference 9.

5.3 TEST RESULTS FOR XSTAR PREPROCESSING

In order to evaluate the performance of XSTAR and of other preprocessing algorithms in a manner which is independent of subsequent intended uses for the preprocessed data (uses which may confound the residual preprocessing error with their own performance limitations), we have measured preprocessing error as the residual error in matching one day's Landsat data to the consecutive day's data, averaged over all pixels in the scene, and have expressed it as a Euclidean distance (root sum square error) in Landsat counts. This performance measure is equivalent to the Euclidean distance error in matching the preprocessed scene means for the two acquisitions. Data flagged by the SCREEN algorithm (garbled data, clouds, snow, dense haze concentrations, cloud shadows, and water) has been excluded from these residual error calculations. Some additional performance measures which elucidate other characteristics of the residual error in matching data from consecutive days have also been examined and are discussed later in this section.

A simple test was performed to estimate the relation between residual Euclidean distance error in preprocessing and loss of recognition accuracy from signature extension, excluding the usual loss of accuracy caused by imperfect training data. For this test, training signatures derived from the Finney Co., Kansas, Intensive Test Site for April 20, 1974, were modified by shifting their mean values in a manner which simulated varying amounts of error in matching the signatures to the data. The direction in which the signature means were shifted was chosen to simulate typical shifts in Landsat signals caused

by increasing or decreasing haze as observed empirically in Landsat II data. Since the data chosen for the test was Landsat I data, the direction of the shift applied to the signature means was adjusted to take into account the calibration differences between Landsat I and Landsat II. The effect of the various simulated errors on the wheat proportion estimate for this scene is plotted in Figure 12. In this figure positive error refers to shifting the signature means in the direction of positive correlation with the Tasselled Cap brightness axis (i.e., increasing haze). (Our experience with the XSTAR preprocessing algorithm has indicated that "positive" and "negative" preprocessing errors are about equally probable.) From Figure 12 we judge that up to three counts Euclidean distance error may be tolerable, while errors in excess of three counts may not be tolerable.

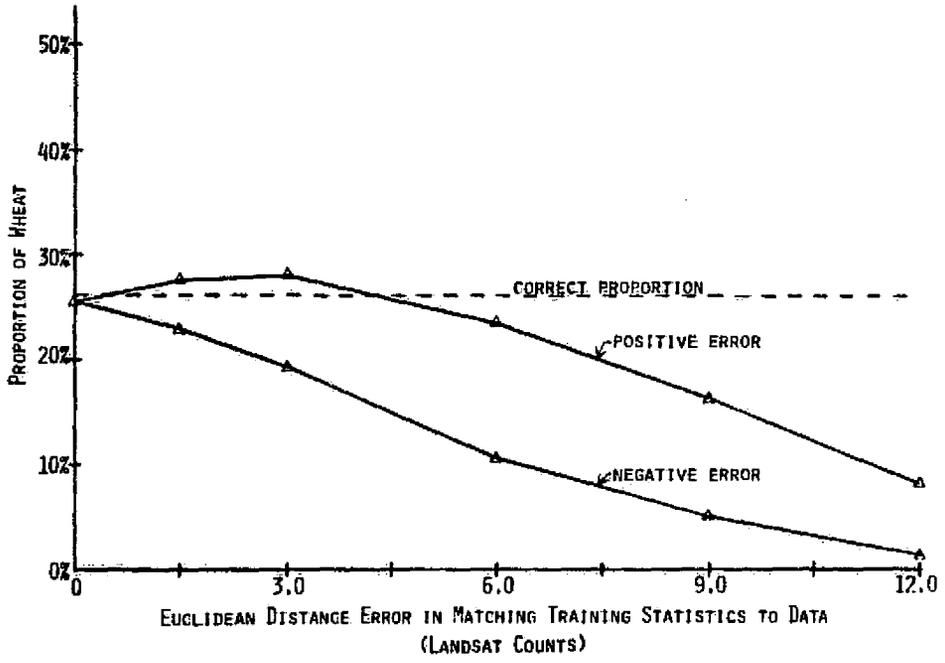


FIGURE 12. ACCURACY OF WHEAT PROPORTION ESTIMATE VS. EUCLIDEAN DISTANCE ERROR IN MATCHING TRAINING STATISTICS TO DATA. Finney Co., Kansas, Intensive Test Site, 20 April 1974 (Threshold = 0.001). Simulated errors represent typical effects of increasing and decreasing haze on training statistics used for signature extension.

To determine the amount of preprocessing error to be expected in matching Landsat data from consecutive days, with and without using the XSTAR algorithm, 58 winter wheat data sets and 33 spring wheat data sets were prepared. Each data set consisted of an eight channel Landsat II LACIE sample segment (117 scan lines, with 196 pixels per line), composed from a consecutive day pair of acquisitions. The data sets were clustered in an unsupervised manner, producing up to 100 eight channel clusters per data set. The cluster mean values and the number of pixels used to generate each cluster were then used in lieu of the individual pixel values for the subsequent processing. This greatly reduced the time and costs involved in any tests using the data. The SCREEN algorithm (Section 4 and Reference 9) was used to eliminate clusters from the test which represented garbled data, clouds, snow, dense haze concentrations, cloud shadows, or water in any acquisition.

The XSTAR preprocessing algorithm was tested on the 91 data sets described above as follows. First, since the XSTAR algorithm was derived for a fixed sun zenith angle ($\sim 39^{\circ}$), and since the data sets had sun zenith angles ranging from 31° to 68°, a cosine correction was applied to each data set to simulate data acquired for a sun zenith angle of 39°. The XSTAR haze diagnostic was then determined independently for each day's data for each consecutive day pair. The haze diagnostic was calculated from the averages of the appropriate cluster mean values, weighted by the number of pixels in each cluster, but excluding those clusters flagged by the SCREEN algorithm on either day. On the average, about 11% of the pixels were edited out from each winter wheat data set by the SCREEN procedure, while for the spring wheat data sets, on the average, 22% of the pixels were edited out. However, the SCREEN procedure edited out all but three of the clusters (containing a total of only 140 pixels) from one of the 33 spring wheat data sets (shown, as it happens, in Figure 5). Since statistical results from this data set would have been of dubious value, it was excluded from the testing.

The results from these tests for XSTAR preprocessed data and for data with no preprocessing (except for a cosine sun zenith angle correction) were sorted in order of increasing magnitude of the Euclidean distance error and are displayed in Figures 13 and 14. Remembering that from interpreting Figure 12 we proposed that 3 counts Euclidean distance error be considered an approximate upper bound for acceptable preprocessing performance, note that for the data sets in Figure 13, while only 16 of the cases with no preprocessing had less than 3 counts error, 31 of the XSTAR preprocessed cases were within this limit. For the spring wheat data sets in Figure 14, XSTAR preprocessing increased the number of cases with less than three counts Euclidean distance error from 10 to 20. Note that for the majority of the test results

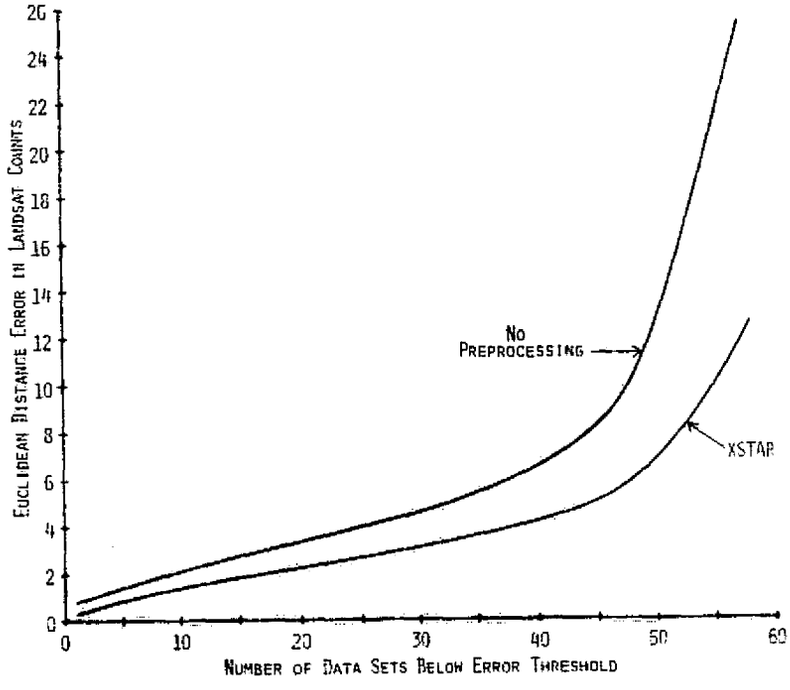


FIGURE 13. SCENE AVERAGE EUCLIDEAN DISTANCE ERROR FROM XSTAR TEST ON 58 CONSECUTIVE DAY WINTER WHEAT DATA SETS (After Cosine Correction for Sun Angle)

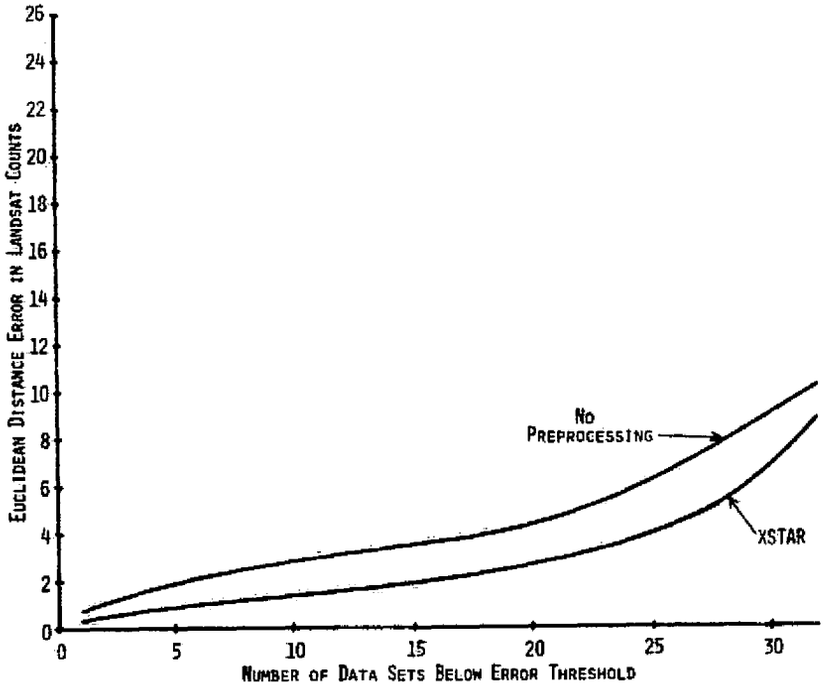


FIGURE 14. SCENE AVERAGE EUCLIDEAN DISTANCE ERROR FROM XSTAR TEST ON 32 CONSECUTIVE DAY SPRING WHEAT DATA SETS (After Cosine Correction for Sun Angle)

shown in Figure 13, the amount of error decreased by approximatley 33% after preprocessing with XSTAR, while for the results shown in Figure 14, the decrease in the error was usually between 33% and 50%. For those cases shown in which the XSTAR preprocessing was not as effective as desired, XSTAR still produced significant improvement compared to using no preprocessing. Those data sets for which XSTAR did least well were those having varied haze within a single acquisition or those with more than 20% of the scene covered by clouds, cloud shadows, or snow. A more thorough screening procedure (e.g., biased in favor of errors of commission) for avoiding the effects of clouds, cloud shadows, and snow on the XSTAR haze diagnostic could possibly improve the results for those cases.

The seemingly better performance of XSTAR on the spring wheat data, relative to the winter wheat data, is presently suspected to be due to the respective amounts of change in view angle. For consecutive day data from Kansas, as in Figure 13, the change in view angle is between 7 and 7.5 degrees, while for consecutive day data from North Dakota and Montana, as in Figure 14, the change in view angle is about 6 degrees. As yet XSTAR does not fully compensate for changes in Landsat signals with view angle.

The results presented above for XSTAR primarily test its ability to compensate for the effects of haze on an average signal. A second order measure of performance would test the accuracy of the multiplicative factors predicted by XSTAR, which affect the correction of signals relative to an average signal. Since the XSTAR multiplicative factors are based on the atmospheric attenuation estimated by ERIM's radiative transfer model, but do not include effects of changing view angle, a test was performed to empirically verify this estimate. In this test the multiplicative factors defined by a pixel by pixel regression (simulated by using cluster means weighted by the number of pixels in each cluster) were analyzed for the 58 winter wheat consecutive day data sets as follows. First, channel by channel averages of the logarithm of each multiplicative factor were computed for data sets with an average logarithm greater than zero and for data sets with an average logarithm less than zero. These averages were then subtracted one from the other in each channel to mostly remove any systematic multiplicative effect correlated with view angle. The four logarithms thus obtained were then rescaled so that their average value (averaging over the four Landsat bands) was unity. This procedure produced four values, derived empirically, which could be compared, band by band, with the a coefficients of the XSTAR algorithm (Equation 21). For a data set requiring a multiplicative factor of 2 (representing a change in optical thickness of ln 2), Table 1 compares these empirically

TABLE 1. ATTENUATION FACTORS IN LANDSAT II DATA DUE TO ATMOSPHERIC TRANSMISSION. (Estimated from 58 Consecutive Day Winter Wheat Data Sets -- Average Simulated Attenuation = 2.0)

	XSTAR Model	Empirical Estimate	Difference
Band 4	2.41	2.49	-3.5%
Band 5	2.06	1.96	5.1%
Band 6	1.88	1.88	0.4%
Band 7	1.71	1.74	-1.8%

estimated multiplicative factors to the XSTAR multiplicative factors. Even for such an extreme case the agreement is quite close, although the procedure used to derive the empirical estimate was at best only approximate.

A second test was performed to determine, on an empirical basis, whether there was any systematic multiplicative effect correlated with view angle (i.e., to quantify those multiplicative effects which had been deliberately excluded in the comparison to the α coefficients of XSTAR in Table 1). This test for systematic view angle effects was performed in two different ways. The first way was to combine the multiplicative factors determined separately for each day's data by the XSTAR algorithm and then to compare these to the multiplicative factors (subject to the α constraint) which matched the data for the two days of each data set in a least squares sense. The average systematic multiplicative effect determined in this manner is shown in Table 2 for both the winter wheat data sets and the spring wheat data sets. Note that the winter wheat data, with the larger change in view angle, exhibits the larger effect. Although this systematic multiplicative effect is nearly an order of magnitude larger than had been anticipated from signature modeling results for winter wheat canopies

TABLE 2. AVERAGE CHANGE IN LANDSAT II SIGNAL CONTRAST UNACCOUNTED FOR BY XSTAR, WHEN APPLIED TO CONSECUTIVE DAY DATA. (Day 1 to Day 2)

Winter Wheat Data Sets (Kansas)	-16% ± 2%
Spring Wheat Data Sets (N. Dakota and Montana)	-10% ± 3%

which were reported in our first contract quarterly progress report [12], the total effect is not enough to be seriously detrimental to the performance of XSTAR.

The second way a test for systematic multiplicative effects with view angle was performed was to use the simulated pixel by pixel regression estimates for the multiplicative factors. The logarithms of these factors were averaged band by band over the 58 winter wheat data sets. Each average logarithm was then converted to a corresponding "average" multiplicative factor by calculating its antilog. Since atmospheric conditions should not be partial to either the first or second day of a consecutive day acquisition, the atmospheric variations should cancel each other in this average, leaving just the systematic multiplicative effects. The results for each Landsat band are shown in Table 3. With the possible exception of band 7, the effect is about the same in each band and is similar in magnitude to the result shown for winter wheat in Table 2.

TABLE 3. AVERAGE CHANGE IN LANDSAT II SIGNAL CONTRAST FOR 58 CONSECUTIVE DAY WINTER WHEAT DATA SETS. (Day 1 to Day 2)

	Percent
Band 4	-22
Band 5	-20
Band 6	-22
Band 7	-17

Assuming that the systematic multiplicative effect is in fact correlated with changes in view angle, the two most probable causes for the effect are (1) changes in the Landsat MSS sensitivity with view angle, or (2) bidirectional variations in the canopy reflectance of average canopies (i.e., not necessarily wheat). To test the first hypothesis, solar calibration data for Landsat II, which had been obtained from personnel at the Goddard Space Flight Center, was analyzed. Since the sun is introduced into the field of view of the Landsat satellite via one of four facets on a deflector mirror, according to the seasonal and shorter term changes in the relative position of the sun and the satellite, one might expect the sun's image to be recorded at different view angles for different acquisitions. Indeed this happens, so that the mean value of the sun at each view angle available can be used to estimate variations in the scanner sensitivity for signal levels near that of the sun. Thus, the solar calibration data (which included 22 acquisitions between 29 April 1975 and 3 August 1976, and provided samples for the entire Landsat field of view) was examined to determine variations in the apparent mean signal value for the sun as a function of view angle. The results of this analysis are presented in Table 4. The data indicated that any

TABLE 4. RESULTS OF ANALYSIS OF 22 LANDSAT II SUN CALIBRATION DATA ACQUISITIONS

	Decompressed Mean Signal Level (in Landsat Counts)	Standard Deviation
Band 4	75.1	3.8%
Band 5	88.2	3.8%
Band 6	79.2	4.7%
Band 7	31.5	2.6%

NOTE: No significant change in signal level with view angle was apparent in Bands 4–7.

changes in the Landsat sensitivity with view angle were probably less than 2%. The standard deviation figures in Table 4 indicate that if variations in the Landsat sensitivity had been the cause of the systematic multiplicative changes observed on consecutive days, the solar calibration results should have made this apparent. Hence, the most probable cause for the systematic multiplicative changes is judged to be bidirectional variations in the canopy reflectance of average canopies.

In passing, we wish to remark that although the sun to earth distance changes throughout the year so that the solar irradiance at the top of the earth's atmosphere varies seasonally by $\pm 3.5\%$, the solar calibration mean values are calculated using only pixels whose instantaneous field of view (IFOV) falls within the solar disk. ence, one would not expect the changes in the sun to earth distance to alter the mean values calculated for the solar calibration procedure. This assumption has been followed in generating the numbers listed in Table 4. However, although the conclusions drawn above from Table 4 would not be affected, we have noted that the solar calibration mean values, when plotted vs. time, appear to exhibit a $\pm 2\%$ variation in Bands 4, 5, and 6, and a $\pm 1\%$ variation in Band 7, which are strongly correlated with the seasonal changes in solar irradiance. This apparent effect could be caused by a rather blurred Landsat IFOV, when viewing the sun (the sun cal optical path differs somewhat from that used when viewing the earth), or by an interaction between the total brightness of the sun and the Landsat calibration procedures (e.g., stray light affecting the signals from the calibration wedge), or by coincidence. More needs to be known about the performance of the Landsat satellites.

5.4 CONCLUSIONS FROM TESTS OF XSTAR PREPROCESSING

In Section 5.3 test results have been presented for XSTAR preprocessing which measure the residual error in matching Landsat II data from consecutive days over 91 separate scenes, representing a wide

range of sun zenith angles, scene characteristics, and atmospheric conditions. Although the XSTAR algorithm is based on a highly simplified model which does not include the effects caused by changes in view angle or background albedo (which are known to be significant), it has nevertheless been significantly effective in reducing the effects of atmospheric haze in Landsat data. For the 91 test cases examined, XSTAR preprocessing, compared to no preprocessing, doubled the number of consecutive day data sets that matched within 3 Landsat counts Euclidean distance (an estimated upper bound on acceptable performance). In all, one half to two thirds of the data sets were brought within 3 Landsat counts of matching, after applying XSTAR, while the remaining data sets (scenes more than 20% covered by clouds, cloud shadows, or snow) were in general significantly improved by XSTAR. Additional experiments, illustrating the effects of XSTAR preprocessing in improving the analysis, interpretation, training value, or classification accuracy (using signature extension) of Landsat data are reported in References 2, 3, and 4. A condensed, detailed programmer's description of the XSTAR algorithm and of its haze diagnostic procedure is presented in Reference 9.

An empirical analysis of the multiplicative factors appropriate for signature extension preprocessing has revealed that although the values of α (Equation 21) chosen to characterize atmospheric attenuation may be reasonably accurate, a significant reduction in apparent scene contrast occurs between the first and second days of a consecutive day Landsat acquisition. This multiplicative effect appears to be related to view angle effects in the bidirectional reflectance of typical crop canopies. This observed effect is one of the reasons for including the multiplicative factor Q (Equation 17) in Equation 18. Recent test results indicate that the performance of the XSTAR algorithm may be improved by applying a small multiplicative "Q" correction to the data together with the sun zenith angle cosine correction before preprocessing with XSTAR. However, a more satisfactory result is likely to be obtained using the XBAR approach discussed in Section 6.

6

THE XBAR SIGNATURE EXTENSION PREPROCESSING ALGORITHM

6.1 DERIVATION

The XBAR signature extension preprocessing algorithm is currently under development and is intended to compensate Landsat data not only for the effects of atmospheric haze, but also for the effects of view angle and background albedo (not attempted in the XSTAR approach) and of sun zenith and azimuth angle (in a more precise manner than the simple cosine correction mentioned in Sections 4 and 5). At present the XBAR algorithm is not intended to compensate for the effects of atmospheric absorption, however a mathematical formulation for such a modification to XBAR (which will not be presented here) has been defined.

The XBAR algorithm is based on the detailed form of the ERIM radiative transfer model as expressed in Equations 2 through 9 (Section 3.2), but with a few more details added. First, since the direct solar irradiance at the top of the atmosphere (E_o, Equation 3) is known to vary seasonally by $\pm 3.5\%$ as the distance from the sun to the earth changes, we have replaced E_o in Equation 3 with the expression from Equation 42, below,

$$E_o = \frac{\overline{E}_o}{D} \tag{42}$$

with

$$D \equiv 1 - .035 \cos\left[2\pi \frac{(\text{Julian date} - 3)}{365.25}\right] \tag{43}$$

The quantity $\overline{E}_o$ in Equation 42 represents the average direct solar irradiance at the top of the atmosphere (averaged over a period of one year), while the quantity "Julian date" in Equation 43 refers to

the Julian date of the data acquisition to be preprocessed. The closest approach of the sun occurs around January 3, hence the square root of D is proportional to the sun-earth distance.

Next, a detailed form has been defined for the factor Q (Equation 17), representing the multiplicative effect of bidirectional reflectance for a typical scene component. A proposed form is

$$Q \equiv 1 + \varepsilon_Q \theta \tag{44}$$

with θ representing the scanner view angle relative to nadir (going from a negative value at the beginning of each scan to a positive value at the end of each scan), and with ε_Q representing a fixed scalar parameter yet to be determined. As a result of future analyses, a more elaborate (and more accurate) formulation for the factor Q may be determined.

Finally, since the ERIM radiative transfer model is based on the assumption of an infinitesimal target surrounded by a uniform background, we have defined a more detailed form for the background albedo, ρ. Note that in practice the effective background albedo for a given target is a spatially weighted average of the reflectances of surrounding materials. This spatial weighting emphasizes the reflectance of nearby materials over that of more distant materials [13]. Since our goal (for the present) is to devise preprocessing techniques which define a single set of multiplicative and additive factors to be applied to a whole scene (usually the size of a LACIE sample segment, containing about 20,000 Landsat pixels), a spatial weighting technique for estimating background albedo is beyond the scope of our present efforts. However, since useful targets in Landsat data are usually larger than one pixel, we may obtain a crude approximation to a spatial weighting technique by using a weighted average of target and average background reflectance. Thus, for the present we define a

scalar weighting factor, ζ (yet to be determined), and let

$$\rho \equiv (1 - \zeta)\,\bar{\rho} + \zeta\rho_t \tag{45}$$

with $\bar{\rho}$ representing the average background reflectance for the scene, and ρ_t representing the target reflectance. The scalar weighting factor ζ would be expected to depend on the average field size and on the optical thickness for each scene, however at present it is included as a simple fixed parameter (yet to be determined) to allow us to study the possible effects of a more precise representation of the influence of the background albedo.

Using Equations 42 and 45, Equation 10 (the condensed form of Equation 2) takes the following form

$$L = \frac{\bar{E}_o \hat{\Lambda}}{4\pi D}\left[\frac{4\mu_o^2\bar{\rho}}{\hat{\Delta}}\left(\hat{K}_o\frac{\rho_t}{\bar{\rho}} + K_1\right) + K_2\right] \tag{46}$$

with

$$\hat{\Lambda} = [\mu_o + (1 - \eta)\tau]^{-1} \tag{47}$$

$$\hat{\Delta} = \frac{1 + 2\,(1 - \rho)(1 - \eta)\tau}{1 - \zeta} \tag{48}$$

$$\hat{K}_o = \frac{K_o + \zeta\,K_1}{1 - \zeta} \tag{49}$$

and the remaining terms of Equations 46 through 48 defined in Equations 4 through 9 and 11 through 13 (Section 3.2).

We also define

$$\hat{G} \equiv \frac{G\bar{E}_o}{4\pi} \tag{50}$$

hence Equation 15 becomes

$$x = \hat{G}\,\frac{\hat{\Lambda}}{D}\left[\frac{4\mu_o^2\,\bar{\rho}}{\hat{\Delta}}\left(\hat{K}_o\,\frac{\rho_t}{\bar{\rho}} + K_1\right) + K_2\right] + \delta \qquad (51)$$

Calculating $\bar{x}$, the average signal level for the scene, and noting that $\bar{\rho}_t \simeq \bar{\rho}$ and that $\hat{\Delta}$ is a weak function of ρ (i.e., $\bar{\hat{\Delta}} \simeq \hat{\Delta}$), we obtain

$$\bar{x} \simeq \hat{G}\,\frac{\hat{\Lambda}}{D}\left[\frac{4\mu_o^2\,\bar{\rho}}{\hat{\Delta}}\,(\hat{K}_o + K_1) + K_2\right] + \delta \qquad (52)$$

or

$$\frac{4\mu_o^2\,\bar{\rho}}{\hat{\Delta}} \simeq \frac{D(\bar{x} - \delta)}{\hat{G}\,\hat{\Lambda}} - K_2 \over \hat{K}_o + K_1 \qquad (53)$$

Equation 53 permits the calculation of the average background albedo, $\bar{\rho}$, from the average signal level, $\bar{x}$, provided that the other quantities in the equation are known. This procedure for estimating the background albedo has led to the name XBAR for the resulting algorithm. We will return to Equation 53 later.

Equations 51 and 52 may be combined to produce

$$x - \bar{x} = \hat{G}\,\frac{\hat{\Lambda}}{D}\,\frac{4\mu_o^2\,\bar{\rho}}{\hat{\Delta}}\,\hat{K}_o\left(\frac{\rho_t}{\bar{\rho}} - 1\right) \qquad (54)$$

Similarly, for the standardized condition, denoted by primes and for which we choose $D' \equiv 1$,

$$x' - \bar{x}' = \hat{G}\,\hat{\Lambda}'\,\frac{4\mu_o'^2\,\bar{\rho}'}{\hat{\Delta}'}\,\hat{K}_o'\left(\frac{\rho_t'}{\bar{\rho}'} - 1\right) \tag{55}$$

Using the definition of the factor Q (Equation 17), Equations 54 and 55 may then be combined, producing

$$x' = DQ\,\frac{\hat{\Lambda}'\mu_o'^2\,\hat{\Delta}\hat{K}_o'}{\hat{\Lambda}\mu_o^2\hat{\Delta}'\hat{K}_o}\,(x - \bar{x}) + \bar{x}' \tag{56}$$

(The reader may note the appearance of an analogy between Equation 56 and Equation 40.)

The next step is to calculate $\bar{x}'$ from known quantities. This is accomplished by writing the equivalent of Equation 52 for the standardized scene (adding primes to the appropriate variables) and by substituting for $\bar{\rho}'$ in this equation, using Equations 17 (the Q factor) and 53. This produces

$$\bar{x}' = DQ\,\frac{\hat{\Lambda}'\mu_o'^2\hat{\Delta}(\hat{K}_o' + K_1')}{\hat{\Lambda}\mu_o^2\hat{\Delta}'(\hat{K}_o + K_1)}\,\bar{x} + \left(1 - DQ\,\frac{\hat{\Lambda}'\mu_o'^2\hat{\Delta}(\hat{K}_o' + K_1')}{\hat{\Lambda}\mu_o^2\hat{\Delta}'(\hat{K}_o + K_1)}\right)\delta$$

$$+ \hat{\Lambda}'\left(K_2' - DQ\,\frac{\mu_o'^2\hat{\Delta}(\hat{K}_o' + K_1')}{\mu_o^2\hat{\Delta}'(\hat{K}_o + K_1)}\,K_2\right)\hat{G} \tag{57}$$

Finally Equations 56 and 57 may be combined to obtain

$$x' = DQ \frac{\hat{\Lambda}'\mu_o'^2\hat{\Delta}\hat{K}_o'}{\hat{\Lambda}\mu_o^2\hat{\Delta}'\hat{K}_o}(x - \bar{x}) + DQ \frac{\hat{\Lambda}'\mu_o'^2\hat{\Delta}(\hat{K}_o' + K_1')}{\hat{\Lambda}\mu_o^2\hat{\Delta}'(\hat{K}_o + K_1)}\bar{x}$$

$$+ \left(1 - DQ \frac{\hat{\Lambda}'\mu_o'^2\hat{\Delta}(\hat{K}_o' + K_1')}{\hat{\Lambda}\mu_o^2\hat{\Delta}'(\hat{K}_o + K_1)}\right)\delta + \hat{\Lambda}'\left(K_2' - DQ \frac{\mu_o'^2\hat{\Delta}(\hat{K}_o' + K_1')}{\mu_o^2\hat{\Delta}'(\hat{K}_o + K_1)} K_2\right)\hat{G} \quad (58)$$

Equation 58 is the XBAR equivalent of Equation 18 (Section 3.2). The XBAR algorithm also uses Equations 20 and 22 (or 19), which relate the aerosol optical thickness to the scalar parameters γ and γ'.

6.2 IMPLEMENTATION OF THE XBAR ALGORITHM

At present the unknown quantities in Equation 58 are $\hat{G}$, δ, γ', γ, ε_Q, and ζ. Although estimates for $\hat{G}$ and δ (needed for each Landsat spectral band) could be obtained from the Landsat prelaunch calibration, we would not expect these estimates to be sufficiently accurate for this application, however the prelaunch values could serve later as a check on our calculations. We do plan to calculate $\hat{G}$ and δ by performing a regression over selected consecutive day Landsat data sets. The form of Equation 58 makes such a regression straightforward, and in fact the procedure has already been programmed and checked out on a computer. The least squares estimates we have calculated so far for $\hat{G}$ and δ have exhibited a strong dependence on our trial values for γ' and ε_Q. The parameter ζ has so far been left set to zero.

Proper estimates for $\hat{G}$ and δ also require the implementation of a suitable haze diagnostic procedure, however in order to utilize a haze diagnostic based upon the XBAR model, one must already have estimates for $\hat{G}$ and δ. For the interim the XSTAR haze diagnostic has been used to estimate γ so that with trial values for γ' and ε_Q we could

at least obtain some preliminary estimates for $\hat{G}$ and δ. Once we obtain some dependable values from these preliminary estimates, an XBAR haze diagnostic can then be used to estimate γ, hopefully leading to still better and more consistent estimates for $\hat{G}$ and δ.

Given trial values for $\hat{G}$, δ, γ', γ, ε_Q, and ζ, we can also calculate the average background albedo, $\bar{\rho}$, from Equation 53. Since the realistic range of values for $\bar{\rho}$ is rather limited, this serves as a very sensitive check on the performance of the XBAR algorithm. By iterating through successive estimates for $\hat{G}$ and δ, which in turn would lead to more accurate haze diagnostics from XBAR, and by monitoring estimates for $\bar{\rho}$, we expect to converge on an operational XBAR implementation in the near future.

6.3 COMMENTS ON THE XBAR ALGORITHM

Once sufficiently accurate estimates for $\hat{G}$, δ, γ', ε_Q, and ζ have been obtained, the application of the XBAR preprocessing algorithm to Landsat data will be very similar to the present application of the XSTAR algorithm, the significant difference being that the XBAR haze diagnostic calculation and the calculation of the XBAR multiplicative and additive correction factors would be more detailed than in the XSTAR correction, and that the corresponding preprocessing would be more accurate. The scattering phase functions ($p(\mu, \phi, -\mu_o, \phi_o)$ and $p(\mu, \phi, \mu_o, \pi + \phi_o)$) required by XBAR would be calculated by interpolating in a table stored in the computer. For this interpolation the Landsat view angle relative to nadir, the latitude of the scene, and the sun zenith and azimuth angles at the time of the data acquisition would have to be known. Since the XBAR calculations would only be done once for each scene ($\sim$20,000 pixels), the increase in cost relative to the XSTAR algorithm would be small.

The XBAR preprocessing algorithm would provide one substantial benefit in addition to the preprocessing of Landsat data -- the definition of a close correspondence between Landsat MSS data and a detailed

radiative transfer model. This would allow a more complete utilization of ground measurements in remote sensing experiments. It would also provide a powerful technique for monitoring the performance of the Landsat MSS system.

PRELIMINARY ANALYSES OF SOIL COLOR EFFECTS IN LANDSAT AGRICULTURAL DATA

In addition to the effects of changing atmospheric haze, changes in soil color or soil condition can also significantly affect Landsat signals not only between scenes, but from field to field as well [14]. Some studies have been underway at ERIM to determine the effect of soil color or soil condition on Landsat agricultural data and to try to develop preprocessing techniques for removing or reducing the confusing effects of soils in a way that would improve signature extension performance. A prerequisite for these studies, however, has been the development of a reliable preprocessing technique for removing the effects of atmospheric haze from Landsat data. For the present the XSTAR signature extension preprocessing algorithm satisfies this requirement.

Some basic insights regarding the effect of soil color on Landsat data have been obtained from signature modeling. H. R. Condit reported measurements of spectral reflectances for a wide variety of soils sampled throughout the continental U.S., and found that more than 93% of the variance of these soil reflectances in the range of wavelengths from 0.32 to 1.0 microns could be represented by a linear combination of a reflectance mean vector and a single displacement vector [15]. Using this linear combination to simulate a dark, a medium, and a bright soil, we have used the Suits canopy model [16,17,18] to simulate the effect of these soils on the Landsat in-band reflectance of emergent wheat canopies, with various canopy densities and leaf orientations [4]. The result is shown in Figure 15, with the four Landsat in-band reflectance coordinates rotated to simulate a plot of green vs. brightness from the Tasselled Cap [8,4]. Note that the line segments in the figure, simulating the effect of changing soil brightness, all seem to point toward a single location in the reflectance data space,

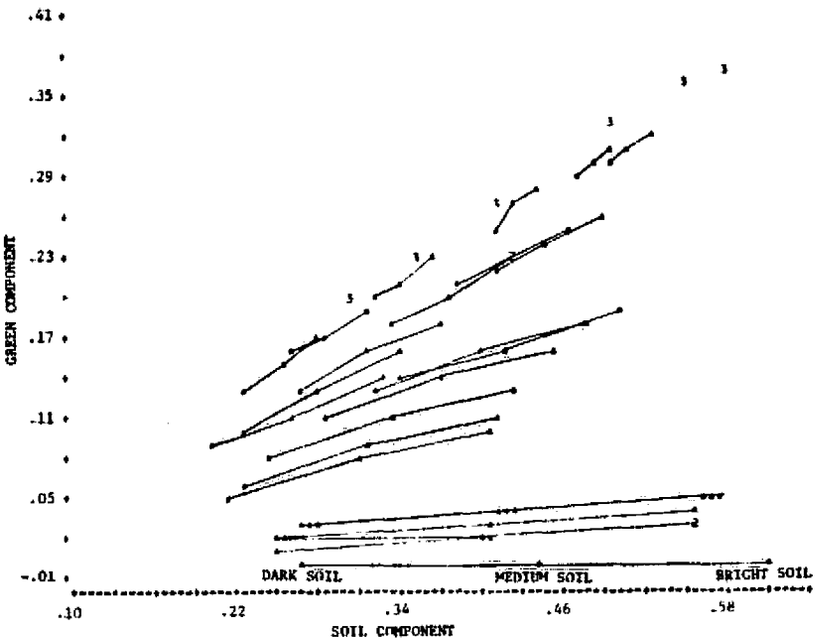


FIGURE 15. SIMULATED EFFECT OF SOIL BRIGHTNESS ON LANDSAT IN-BAND EMERGENT WHEAT CANOPY REFLECTANCE

near zero reflectance. This suggests that a ratio of Tasselled Cap greenness to brightness or of Landsat band 7 to 5 may eliminate much of the variability due to soil brightness. We also suspect that the amount of important information about wheat canopies which is confounded with these soil brightness variations is relatively small compared to the information contained in either of these ratios. (An experiment is planned to test the effectiveness of these signal ratios as a signature extension preprocessing technique.)

Another possible approach to minimizing the effects of soil variability in Landsat data is to characterize the temporal development of vegetation on soil with sufficient accuracy or detail so that a data acquisition early in the growing season may be used to estimate the soil brightness or color for each field of interest and so that the

future effects of the soil on the Landsat signal from each field can then be reliably predicted. This would in effect be a preprocessing technique to remove or to reduce the effects of soil brightness or soil color in Landsat data, which would require a calibration step to be performed early in the growing season. We are presently analyzing Landsat data and the available supporting ground observations to explore this approach to soil brightness preprocessing. Other simpler approaches to this preprocessing problem are expected to come to light as by-products of this line of research.

Our analyses of soil effects in Landsat data have been concentrated on data from the LACIE intensive test sites (ITS's), since this is virtually the only available data with supporting ground observations in any sufficient detail. These ground observations were planned to correspond reasonably well with the times of the Landsat overpasses, however observations during the fall of the year, when soil effects on emerging winter wheat can best be studied, exist only for a very few of the ITS's. Of these, only one (Finney Co. ITS, Kansas, 1975-76 crop year) has a sufficient number of acquisitions to be significantly useful. Hence, our analyses have been hampered by the limited amount of ground information available to support this particular study. In spite of these difficulties, however, some useful insights are emerging.

First, using XSTAR preprocessed Landsat data from four of the ITS's (Finney and Saline Co.'s, Kansas, Randall Co., Texas, and Whitman Co. (2), Washington), we have examined data from fields (mostly fallow) confirmed by ground observations to be bare with minimal weed growth. Pooling this data together, we have found that more than 90% of the variance observed correlates with a single axial direction in the Landsat data space. This axial direction is angled slightly from the standardized orientation of the Tasselled Cap greenness-brightness hyperplane (Section 3.3), presumably due to the rotational effects of haze on the Landsat data distribution (Section 5.2). However, when

this principal axis of soil variation is projected on to the greenness-brightness plane, the deviation from the Tasselled Cap brightness direction is less than 1°. This indicates that the Landsat II Tasselled Cap axes, which were specially oriented to aid our haze correction efforts, are probably well oriented to suit the phenological interpretations of the Tasselled Cap as well. An analysis of the supporting ground observations for the bare fields, however, indicates that the first principal axis of the soil variability correlates not so much with soil reflectance or surface moisture (which were observed to have mostly random effects) as with field operations (e.g., whether the field had been disked or plowed, and whether stubble was present). For instance, we have found that for dry bare ground in our sample (mean brightness $\sim$87 counts, σ $\sim$7 counts), disking or plowing decreased the brightness by about 13 counts and increased σ to about 11 counts. Fields with standing stubble, on the other hand, (similar in appearance to disked or plowed bare fields), increased in brightness to around 88 counts with a large σ ($\sim$19 counts) when they were disked. Plowed stubble, however, appeared similar to worked bare soil. Two burned fields were 45 counts darker than bare unworked soil. Hence, it appears that the major driving factors affecting soil brightness in this data are field operations which affect the texture of the soil surface and the amount of stubble present.

Analyses are now underway to characterize the stability through the early growing season of the soil appearance for typical winter wheat fields. These analyses have established that a threshold of zero in Tasselled Cap greenness is approximately the minimum detectable level for emergent wheat. However, more significant results are expected in the near future.

8

CONCLUSIONS AND RECOMMENDATIONS

The preceding sections summarize our recent progress in developing preprocessing techniques to compensate Landsat MSS data for physical effects without using ground observations. We believe that some significant gains have been achieved in haze compensation with the XSTAR and SCREEN algorithms. However, still more improvement is desirable, and in fact is expected in the near future (perhaps a further reduction in preprocessing error by a factor of 2) from our development of the XBAR algorithm.

Signature extension preprocessing algorithms which are based on our understanding of physical effects in MSS data provide many obvious benefits, for example they

1. Allow training statistics to be derived from more than one region within a partition to provide more complete and representative training information
2. Enable those statistics to be applied usefully over more extensive areas
3. Remove the need for cluster matching algorithms, which are prone to failure whenever the scenes compared are not nearly equivalent subsets of the data distribution to be expected within a partition
4. Provide a stable data base for studying and developing more advanced uses or interpretations of MSS data.

A sufficiently precise preprocessing algorithm (such as XBAR), however, can provide some additional valuable benefits:

5. Establish a calibration for the MSS data such that predictions from theoretical models may be directly compared with empirical observations, not only qualitatively, but quantitatively

6. Provide a means to more closely monitor the performance of an MSS system.

To aid these developments, more detailed information is needed about the performance of the Landsat satellites (to help explain the rotation of the Landsat data hyperplane (Section 5.2), to understand the unexpected influence of the changing sun-earth distance on the solar calibration data (Section 5.3), and more generally to estimate the stability of the calibration of Landsat data).

The development of the XSTAR, XBAR, and SCREEN algorithms has required the extensive use of empirical data (acquired by the Landsat satellites). Some of this development effort will need to be repeated before these techniques can be applied to Thematic Mapper data or even Landsat C data (or other scanner data). Some of the future signature extension research effort, therefore, should be devoted to generalizing and streamlining the adjustment techniques for these algorithms, so that they may be adapted expeditiously to other uses.

Two fortuitous circumstances with respect to Landsat data have made the task of developing preprocessing techniques to standardize physical effects less difficult than it might have been. The most important of these is the occurrence of areas of overlap in the ground swath covered by Landsat on consecutive days. This has allowed consecutive day data acquisitions to be used for adjusting and testing our algorithms. The second important circumstance is the occurrence of "redundant" information in the Landsat bands 4 through 7. Having only four spectral bands to work with, we have found the existence of this apparent redundancy to be crucial to the development of our haze diagnostic procedures. With a greater number of spectral bands, particularly in the visible portion of the spectrum, this "redundancy" may not be as necessary. This needs to be investigated. Planners and designers of future satellite remote sensing systems should be aware of the importance of the above considerations.

On the other hand, our research into the effects of soil color or soil condition on Landsat data has been hampered by a poverty of detailed ground observations, correlated with Landsat overpasses, during the portion of the winter wheat growing season when soils are most distinguishable. Future field measurement programs should attempt to alleviate this deficiency.

Although we have been significantly successful in compensating Landsat agricultural data for the effects of atmospheric haze, we cannot guarantee that these same preprocessing techniques, without adjustments, will work as effectively in non-agricultural applications. We therefore recommend that these algorithms be tested on non-agricultural Landsat data.

REFERENCES

1. Lambeck, P. F. and D. P. Rice, Signature Extension Using Transformed Cluster Statistics and Related Techniques, ERIM 122700-70-F, Environmental Research Institute of Michigan, Ann Arbor, Mich., May 1976.

2. Kauth, R. J. and W. Richardson, Procedure B: A Multisegment Training Selection and Proportion Estimation Procedure for Processing Landsat Agricultural Data, ERIM 122700-31-F, Environmental Research Institute of Michigan, Ann Arbor, Mich., November 1977.

3. Cicone, R. C., J. L. Stinson, and R. J. Balon, An Evaluation of the Signature Extension Approach to Large Area Crop Inventories Utilizing Space Image Data, ERIM 122700-33-F, Environmental Research Institute of Michigan, Ann Arbor, Mich., November 1977.

4. Malila, W. A. and J. M. Gleason, Investigations of Spectral Separability of Small Grains, Early Season Wheat Detection, and Multicrop Inventory Planning, ERIM 122700-34-F, Environmental Research Institute of Michigan, Ann Arbor, Mich., November 1977.

5. Malila, W. A., R. B. Crane, and R. E. Turner, Information Extraction Techniques for Multispectral Scanner Data, Report No. WRL 31650-74-T, Willow Run Laboratories, The University of Michigan, Ann Arbor, Mich., June 1972, Appendix I.

6. Turner, R. E., Investigation of Earth's Albedo Using Skylab Data, ERIM 102200-20-F, Environmental Research Institute of Michigan, Ann Arbor, Mich., January 1976, Appendices I and II.

7. Malila, W. A., J. M. Gleason, and R. C. Cicone, Atmospheric Modeling Related to Thematic Mapper Scan Geometry, ERIM 119300-5-F, Environmental Research Institute of Michigan, Ann Arbor, Mich., April 1976.

8. Kauth, R. J. and G. S. Thomas, "The Tasselled Cap -- A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by Landsat", Symposium Proceedings on Machine Processing of Remotely Sensed Data, Purdue University, W. Lafayette, Indiana, June 29 - July 1, 1976.

9. Lambeck, P. F., "Revised Implementation of the XSTAR Haze Correction Algorithm and Associated Preprocessing Steps for Landsat Data", ERIM Memo Number IS-PFL-1916, November 1, 1977.

10. Wiegand, Craig L., et al., Soil, Water, and Vegetation Conditions in South Texas, Agricultural Research Service, USDA, Weslaco, Texas, June 1977.

11. Lambeck, P. F., "Implementation of the XSTAR Haze Correction Algorithm and Associated Preprocessing Steps for Landsat Data", ERIM Memo Number IS-PFL-1272, 18 March 1977.

12. Lambeck, P., W. Malila, and J. Gleason, "Signature Extension Preprocessing Algorithm Development", 1st Quarterly Report, Task 2, "Analysis of Scanner Data for Crop Inventory", Report 122700-5-L, Environmental Research Institute of Michigan, Ann Arbor, Mich., September 1976.

13. Turner, Robert E., Atmospheric Effects in Multispectral Remote Sensor Data, ERIM 109600-15-F, Environmental Research Institute of Michigan, Ann Arbor, Mich., May 1975.

14. Colwell, John E. and Gwynn H. Suits, Yield Prediction by Analysis of Multispectral Scanner Data, ERIM 109600-17-F, Environmental Research Institute of Michigan, Ann Arbor, Mich., May 1975, Appendix I.

15. Condit, H. R., "The Spectral Reflectance of American Soils", Photogrammetric Engineering, Vol. 36, No. 9, 1970, pp. 955-966.

16. Suits, Gwynn H., "The Calculation of the Directional Reflectance of a Vegetative Canopy", Remote Sensing of Environment, Vol. 2, No. 2, 1972, pp. 117-125.

17. Suits, Gwynn H., "The Cause of Azimuthal Variations in Directional Reflectance of Vegetative Canopies", Remote Sensing of Environment, Vol. 2, No. 3, 1972, pp. 175-182.

18. Suits, Gwynn H. and G. R. Safir, "Verification of a Reflectance Model for Mature Corn with Applications to Corn Blight Detection", Remote Sensing of Environment, Vol. 2, No. 3, 1972, pp. 183-192.