Reprinted from

Ninth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Natural Resources Evaluation

June 21-23, 1983

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

EFFECTS OF PREPROCESSING LANDSAT MSS DATA ON DERIVED FEATURES

T.M. PARRIS, R.C. CICONE

Environmental Research Institute
of Michigan
Ann Arbor, Michigan

I. ABSTRACT

Important to the use of multitemporal Landsat MSS data for earth resources monitoring, such as agricultural inventories, is the ability to minimize the effects of varying atmospheric and satellite viewing conditions, while extracting physically meaningful features from the data. In general, the approaches to the preprocessing problem have been derived from either physical or statistical models. This paper compares three proposed algorithms; XSTAR haze correction, Color Normalization, and Multiple Acquisition Mean Level Adjustment. These techniques represent physical, statistical, and hybrid physical-statistical models, respectively. The comparisons are made in the context of three feature extraction techniques; the Tasseled Cap, the Cate Color Cube, and Normalized Difference.

II. INTRODUCTION

The launch of Landsat 1 in 1972 provided the remote sensing community with a source for multispectral, multitemporal data and has enhanced our ability to monitor the earth's resources. Multitemporal data has especially enabled the tracking of cyclical vegetative growth patterns and crop conditions of the world's agricultural lands.[1] The Large Area Crop Inventory Experiment (LACIE, 1975-1977) demonstrated accurate large area yield estimates could be derived from Landsat data based on analyst-driven pattern recognition techniques and yield models.[2] Recently, the Agricultural and Resources Inventory Surveys through Aerospace Remote Sensing (AgRISTARS, 1980-present) program's Inventory Technology Development project (ITD) has aimed at reducing the cost/benefit ratio of the data analysis and information extraction process through increased automation. Many of the techniques proposed for these projects preprocess the satellite data in some way to minimize the effects of extraneous conditions such as sensor calibration, sun angle, viewing angle, and atmospheric haze.

The preprocessing problem is formally represented by the relation between the reflectance, ρ, of a target and the radiance, L, of that target as observed by a satellite:

$$L = \frac{E(\tau)e^{-\frac{\tau}{\mu}}}{\pi}\rho+L_p$$

where $E(\tau)$ is the sum of direct and diffuse irradiance on the target, τ is the optical thickness of the atmosphere, μ is the cosine of the viewing angle relative to nadir, and L_p is the path radiance due to scattering in the atmosphere.[3] The approaches to this problem generally fall into one of two categories: those based on physical models, and those based on statistical models.

The XSTAR[4] and ATCOR[5] haze correction algorithms are representative of the physical approaches to preprocessing. Each of these methods compute a haze diagnostic from the satellite data and then use it in simplified radiative transfer models.

Cluster matching algorithms, such as CROP-A[6] and MLEST[7], use statistical techniques in computing a gain-offset transformation that will match the signatures in a recognition segment to those of a training segment. Another variety of statistical preprocessing is represented by the Mean Level Adjustment (MLA)[8] and Color Normalization (CN)[9] algorithms. These algorithms respectively compute additive and multiplicative corrections based on the scene mean.

III. EXPERIMENTAL DESCRIPTION

Two experiments were performed with the goal of comparing the effects of preprocessing algorithms on derived features. The first examines effects of preprocessing on derived features. The second examines the effects of scene content on preprocessing. For the purposes of these experiments, three preprocessing algorithms were chosen for analysis; XSTAR haze correction*, CN, and Multiple Acquisition Mean Level Adjustment (MAMLA).[10] XSTAR and CN represent physical and statistical techniques, whereas MAMLA is a hybrid technique that combines features found in both XSTAR and CN. The effects of each algorithm were observed in the context of three representative feature extraction techniques: the Tasseled Cap Transformation[11], the Cate Invariant Color Transformation[12], and Normalized Difference.[13]

A. DESCRIPTION OF THE ALGORITHMS

Mathematical descriptions of the preprocessing and feature extraction techniques are presented in addenda. A summary of the intuition behind each technique follows.

Preprocessing Algorithms. XSTAR - The XSTAR haze correction algorithm is an outgrowth of a simplified version of the ERIM Radiative Transfer Model.[14] The algorithm is unique in that it derives a haze diagnostic from the satellite data itself and then uses it as a parameter in an atmospheric model. The Yellowness feature of the Tasseled Cap space is used as this haze diagnostic from which a measure of relative optical thickness is computed. The algorithm also includes a cosine correction for sun angle and a gain-offset correction for sensor calibration.

There are two implementations of the XSTAR algorithm. Global XSTAR computes a mean haze diagnostic for an entire scene, whereas Spatially Varying XSTAR computes diagnostics for each non-overlapping 5x5 pixel region, and then smoothes them with a Gaussian filter.[15] In theory, Spatially Varying XSTAR should provide better results when the haze level is not uniform over the entire scene.

Color Normalization - Many of the extraneous effects that one attempts to remove with a preprocessing scheme have been observed to be multiplicative in nature. Hence, simply dividing a group of pixels by their mean should minimize these effects. This technique has been observed to produce desirable results under certain conditions.

*The sun angle and satellite correction in the XSTAR algorithm was also analyzed independently, but due to its similarity to MAMLA, it is not discussed in this paper.

Multiple Acquisition Mean Level Adjustment - This algorithm attempts to combine the intuitive simplicity of the CN algorithm with the sun angle correction and satellite calibration features of the XSTAR approach. The scene means for each acquisition of a segment are adjusted for sun angle and sensor and then combined to form a segment mean. The segment mean is then inversely corrected for sun angle and sensor for each acquisition. The pixel values are subsequently divided by the acquisition specific segment mean.

Feature Extraction. The Tasseled Cap - The Tasseled Cap Transform is an affine transformation that captures the majority of agricultural information in two features known as Greenness and Brightness. Greenness is interpreted to correspond to crop growth and development, whereas Brightness is perceived to correspond to soil color and albedo. A third feature, Yellowness, is used as a haze diagnostic in the XSTAR algorithm. The remaining feature, Nonesuch, is often discarded since it has not been found to contain agriculturally useful information.

The Cate Color Cube - The Cate Invariant Color Transform is a non-linear transformation based on a modified cylindrical coordinate system. The extracted features, termed Hue, Value, and Chroma, are interpreted as relating to the color characteristics of imagery produced by a film generator using MSS Bands 4, 5 and 7.

Normalized Difference - Normalized Difference is interpreted to be a measure of crop development. Similar to the Hue feature of the Cate Color Cube, Normalized Difference is non-linear and describes an angular displacement of data in the two space of MSS Bands 7 and 5.

B. DESCRIPTION OF THE DATA AND PROCESSING

In the first experiment, data collected on 14 passes of Landsats 2 and 3 over Montgomery Co., Indiana (Segment 127) during the 1978 growing season was processed under each of the three preprocessing schemes.† The resulting data was then used as input to each of the three feature extraction techniques. Ground truth information was then used to stratify the data by crop type and temporal profiles of corn, soybean, and trees were produced for each of the resulting features.

†The spatially varying implementation of XSTAR was used.

Segment 127/78 was chosen because several agricultural situations are represented (all soil, all vegetation, and mixed soil and vegetation). Additionally, most dates are very clear, with the exception of Day 197 which shows indications of spatially varying haze.

The second experiment was designed to test the stability with respect to scene content of the preprocessing algorithms. This was done by extracting mean profiles for corn, soy, and trees from the four MSS bands of Segment 127/78 (described above). Five hypothetical scenes containing varying amounts of each crop were then processed under each technique.* The proportions used are displayed in the following chart:

Scene	% Corn	% Soy	% Trees
1	63	32	5
2	80	10	10
3	10	80	10
4	10	10	80
5	33.3	33.3	33.3

IV. RESULTS

A. THE EFFECT OF PREPROCESSING IN MSS SPACE

Figure 1 shows composite scatterplots of Band 7 vs. Band 5 for all 14 acquisitions of the resulting data. In Figure 1a the data along the diagonal primarily correspond to bare soil whereas the data along Band 7 for which Band 5 values are high represent various stages of vegetative development. High values along Band 7 for which Band 5 values are also high are haze affected pixels of Acquisition 197. Figures 1b, 1c, and 1d show the same data after processing by the XSTAR, CN, and MAMLA algorithms, respectively. Note that XSTAR and MAMLA maintain the general shape of the data while it is compressed to compensate for sun angles that ranged from 21° to 59°. In addition, XSTAR has moved those pixels attributed to haze closer to the bulk of the unaffected data. CN behaved differently by compressing the data towards a central point and significantly changing its general shape.

The different behavior of CN can be understood by considering the example presented in Figure 2. In this figure the envelopes of three hypothetical acquisitions in various agricultural situations (all soil, all vegetation, mixed soil and vegetation) are shown before and after CN. One can see that even though these means represent three distinct points in raw data space, they are superimposed in the CN preprocessed space. This is due to the fact that CN considers each acquisition out of context from a physical model or the other acquisitions. MAMLA does not perform this way because it considers each acquisition in context of the others and XSTAR does not perform this way because it works within the context of a physical model.

*Due to the loss of spatial information, global XSTAR was applied to each crop stratum.

B. THE EFFECTS OF PREPROCESSING IN FEATURE SPACE

The most important effect preprocessing has on the features is the change of interpretation that must accompany CN. This may be most easily observed by examining Normalized Difference profiles of trees (Figure 3). The different profiles produced by CN make sense when one realizes that they should be interpreted as a measure of relative crop development rather than absolute measures. Trees start off greener than corn and soybeans, but as corn and soybeans develop, even though the absolute greenness of trees does not change significantly, its relative greenness drops with respect to corn and soybeans. Hence CN produces an apparently inverted profile. This effect, when combined with the non-linearity of the Hue feature produces separation between corn and soybeans that is not available with the other preprocessing methods.

In general, XSTAR and MAMLA produce smoother versions of the raw data profiles without significantly changing their general shape. This is most evident with the Greenness profiles of corn (Figure 4). An exception is the spike that MAMLA introduces in the Hue profiles of corn and soybeans (Figure 5).

The effect of sun angle is most clearly seen in the Brightness feature (Figure 6). CN appears to correct for this fairly well, whereas XSTAR and MAMLA appear to be less effective with the extremely low sun angles at the end of the season. This may be due to the inadequacy of the Lambertian assumption at low sun angles.

C. THE EFFECT OF SCENE CONTENT ON PREPROCESSED FEATURES

Figure 7 shows the sensitivity envelopes of Normalized Difference, Brightness, and Hue. Large envelopes are undesirable since the crop profiles are constant regardless of scene content.

The XSTAR profiles are unchanged by scene content. However, XSTAR, by using Yellowness as a haze diagnostic, will be sensitive to any change in Yellowness due to scene content. This is not expected in vegetative applications but is in others. CN shows a significant amount of sensitivity to scene content in all features. This is particularly true at the height of the growing season. MAMLA reduces this problem significantly but still shows a small degree of sensitivity. It should be noted that only the proportions of three crops were changed. If other crops were introduced, one would expect to see a wider degree of variation in the profiles.

V. CONCLUSIONS

Preprocessing techniques, such as XSTAR, that are based on physical models have the desirable feature of a high degree of stability. This stability is only as good as the independence of the haze diagnostic from agriculturally relevant information (which appears to be a reasonable assumption) and the validity of the model employed. Statistical models, such as CN will show a degree of sensitivity to scene content due to their dependence on data dispersion. However, in cases when a certain amount of information is known about a particular scene, the relational aspect of the features may be desirable. Statistical techniques that use larger sample spaces and orient themselves within a physical model, such as MAMLA, appear to show promise in reducing the amount of content sensitivity while maintaining the intuitional simplicity of purely statistical techniques.

REFERENCES

1. Johnson, C.W. & V.B. Coleman, Semi-Automatic Crop Inventory from Sequential ERTS-1 Imagery, Symp. on Significant Results Obtained from the Earth Resources Technology Satellite-1, NASA/GSFC, New Carrollton, MD, March 1973, pp. 11-15.

2. MacDonald, R.B. & F.G. Hall, LACIE: An Experiment in Global Crop Forecasting, Proc. of Plenary Session, LACIE Symp., JSC-14551, NASA/JSC, Houston, TX, Oct. 1978, pp. 17-48.

3. Malila, W.A., R.B. Crane & R.E. Turner, Information Extraction Techniques for Multispectral Scanner Data, Report WRL 31650-74-T, WRL, Univ. of Mich., Ann Arbor, MI, June 1972, App. I.

4. Lambeck, P.F., Signature Extension Preprocessing for Landsat MSS Data, Final Report 122700-32-F, ERIM, Ann Arbor, MI, Nov. 1977.

5. Lambeck, P.F. & J.F. Potter, Compensation for Atmospheric Effects in Landsat Data, Proc. of the Tech. Sessions, LACIE Symp., JSC-16015, NASA/JSC, Houston, TX, Oct. 1978, pp. 723-738.

6. Lambeck, P.F. & D.P. Rice, Signature Extension Using Transformed Cluster Statistics and Related Techniques, Final Report 109600-70-F, ERIM, Ann Arbor, MI, May 1976.

7. Thadani, S.G., The Maximum Likelihood Estimation of Signature Transformation (MLEST) Algorithm, Symp. Proc. on Machine Proc. of Remotely Sensed Data, LARS/Purdue Univ., W. Lafayette, IN, June 1977, IEEE Cat. No. 77CH1218-7 MPRSA, pp. 66-74.

8. Bizzell, R.M. et al, Crop Identification Technology Assessment for Remote Sensing (CITARS), Vol. 1, Task Design Plan, JSC-09384, NASA/JSC, Houston, TX, Feb. 1975, p. 36.

9. Cate, R.B., D.E. Phinney et al, Interpretation of Landsat Digital Data Using a Cubic Color Model Based on Relative Energies, NASA Rep. SR-LO-00418, LEMSCO, Houston, TX, Feb. 1980.

10. Parris, T.M., Mean Level Adjustment, ERIM Memo No. AD-TP-82-1260, Aug. 1982.

11. Kauth, R.J. & G.S. Thomas, The Tasseled Cap--A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by Landsat, Symp. Proc. on Machine Proc. of Remotely Sensed Data, Purdue Univ., W. Lafayette, IN, June 1976.

12. Lambeck, P.F. & J.F. Potter, Compensation for Atmospheric Effects in Landsat Data, Proc. of the Tech. Sessions, LACIE Symp., JSC-16015, NASA/JSC, Houston, TX, Oct. 1978, pp. 723-738.

13. Rouse, J.W., R.H. Haas, J.A. Schell, & D.W. Deering, Monitoring Vegetation Systems in the Great Plains with ERTS, Third ERTS Symp., NASA SP-351, I:309-313, 1973.

14. Turner, R.E., Atmospheric Effects in Multispectral Remote Sensor Data, Report 109600-15-F, ERIM, Ann Arbor, MI, May 1975.

15. Lambeck, P.F., Spatially Varying XSTAR Haze Correction, New Technology Report, ERIM, Ann Arbor, MI, March 1978.

ADDENDUM-MATHEMATICAL DESCRIPTIONS OF THE PREPROCESSING ALGORITHMS

1. The XSTAR Haze Correction Algorithm

a) Satellite calibration and sun angle correction

$$x_i = \frac{\cos\theta_o}{\cos\theta} \cdot [A_{i,s} \cdot L_i + B_{i,s}] \quad , \text{ where}$$

$\theta_o = 39^o$

θ is the sun azimuth

L_i is the Landsat pixel value for the i^{th} channel

s is the satellite number

A is a matrix of gain calibrations for each channel of each satellite

B is a matrix of offset calibrations for each channel of each satellite.

b) Computation of relative optical thickness

$$a = \sum_{i=1}^{4} \alpha_i^2 (x_i - x_i^*) \hat{y}_i \quad , \quad b = \sum_{i=1}^{4} \alpha_i (x_i - x_i^*) \hat{y}_i \quad ,$$

$$c = \sum_{i=1}^{4} (x_i \hat{y}_i) - y^*$$

$$\gamma = -\frac{b}{a}[1-(1-\frac{2ac}{b^2})^{\frac{1}{2}}], \quad \frac{2ac}{b^2}<1 \Rightarrow \gamma = -\frac{b}{a}$$

where:

$\vec{x}^*$ is an empirically defined point of all haze

$\vec{y}$ is the direction of the Yellowness axis

y^* is an empirically defined normal haze condition

γ is a measure of relative optical thickness

c) Computation of the haze corrected pixel value

$$\vec{L}' = e^{\vec{\alpha}\gamma}\vec{x} + (1-e^{\alpha\gamma})\vec{x}^*$$

where:

$\vec{L}'$ is the corrected Landsat vector

$\vec{\alpha}$ represents the relative aerosol optical thickness for each of the four wavelengths

2. Mean Level Adjustment

$$L_i' = \frac{L_i}{\bar{L}_i}$$

where:

L_i is the Landsat pixel value for the i^{th} channel

$\bar{L}_i$ is the mean signal in the i^{th} channel for the acquisition

L_i' is the corrected Landsat pixel value for the i^{th} channel

3. Multiple Acquisition Mean Level Adjustment

a) Calculate the channel means for each acquisition

$\bar{L}_{ij}$ is the i^{th} channel mean for the j^{th} acquisition.

b) Perform Satellite Calibration and Sun Angle Corrections on the Acquisition Means

$$x_{ij} = \frac{\cos\theta_o}{\cos\theta_j} \cdot [A_{i,s_j} \cdot \bar{L}_{ij} + B_{i,s_j}]$$

where:

θ_o is 39^o

θ_j is the sun azimuth angle for the j^{th} acquisition

s_j is the satellite number for the j^{th} acquisition

A is a matrix of gain calibrations for each channel of each satellite

B is a matrix of offset calibrations for each channel of each satellite

Note the similarity between this step and Step a) of the XSTAR algorithm.

c) Calculate the segment mean

$$x_i' = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$$

where: n is the number of acquisitions.

d) Calculate the acquisition specific segment means

$$x_{ij}'' = \frac{\frac{\cos\theta_j}{\cos\theta_o} \cdot x_i' - B_{i,s_j}}{A_{i,s_j}}$$

Note that this is inverse sun angle correction satellite calibration.

e) Perform mean level adjustment with respect to the acquisition specific segment means

$$L_{ij}' = \frac{L_{ij}}{x_{ij}''}$$

where: l_{ij} is a pixel value from the i^{th} channel of the j^{th} acquisition

l_{ij}' is the corrected pixel value.

ADDENDUM--MATHEMATICAL DESCRIPTION OF THE FEATURE EXTRACTION TECHNIQUES

1. Tasseled Cap Transformation

$$\bar{t} = T\bar{L}$$

where: $\bar{L}$ is a Landsat pixel vector
T is an orthogonal matrix with rows corresponding to Brightness, Greenness, Yellowness & Nonesuch
$\bar{t}$ is the Tasseled Cap pixel vector

2. Cate Color Cube

$$\bar{c} = C\bar{L} \quad , \quad \text{Hue} = \text{Tan}^{-1}\left(\frac{c_2}{c_1}\right)$$

$$\text{Value} = \frac{1}{\sqrt{3}} c_3 \quad , \quad \text{Chroma} = \left(c_1^2 + c_2^2\right)^{1/2}$$

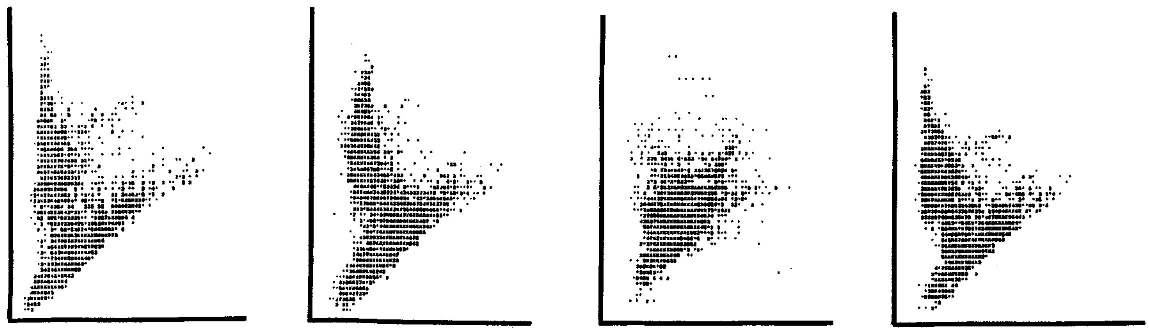
where: $\bar{L}$ is a Landsat pixel vector
C is an orthogonal matrix
$\bar{c}$ is the rotated pixel vector.

3. Normalized Difference

$$N = \frac{L_4 - L_2}{L_4 + L_2}$$

where: $\bar{L}$ is a Landsat pixel vector
N is the normalized difference.

Note: This work was sponsored under Contract NAS9-16538 by the U.S. National Aeronautics and Space Administration, NASA Johnson Space Center, Houston, Texas 77058.



(a) Raw Data (b) XSTAR (c) CN (d) MAMLA

Figure 1. The Effect of Preprocessing in MSS Space. Segment 127 (1978) All Acquisitions.

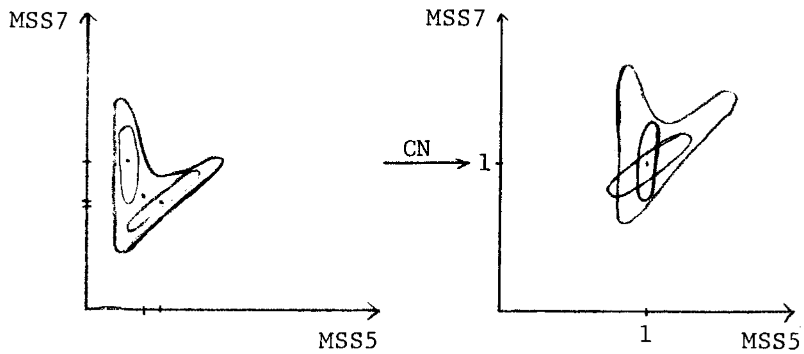


Figure 2. Hypothetical Application of CN on Three Acquisitions in Varying Stages of Vegetative Development.

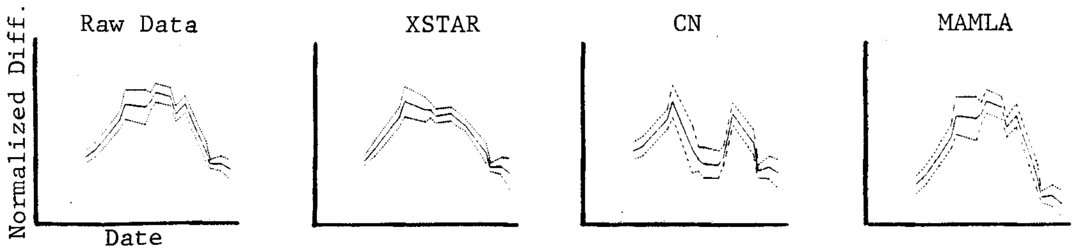


Figure 3. The Effects of Preprocessing on the Normalized Difference Profiles of Trees.

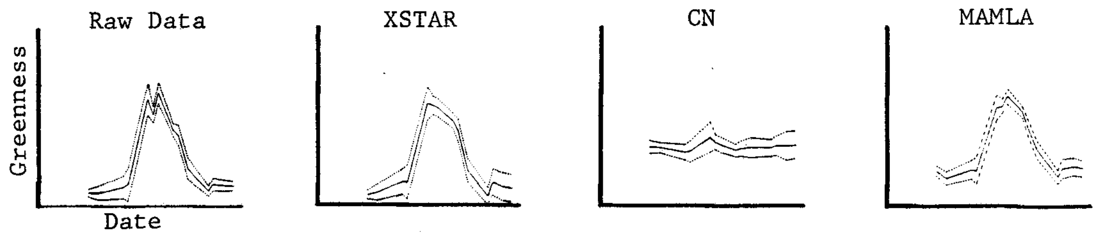


Figure 4. The Effects of Preprocessing on the Greenness Profiles of Corn.

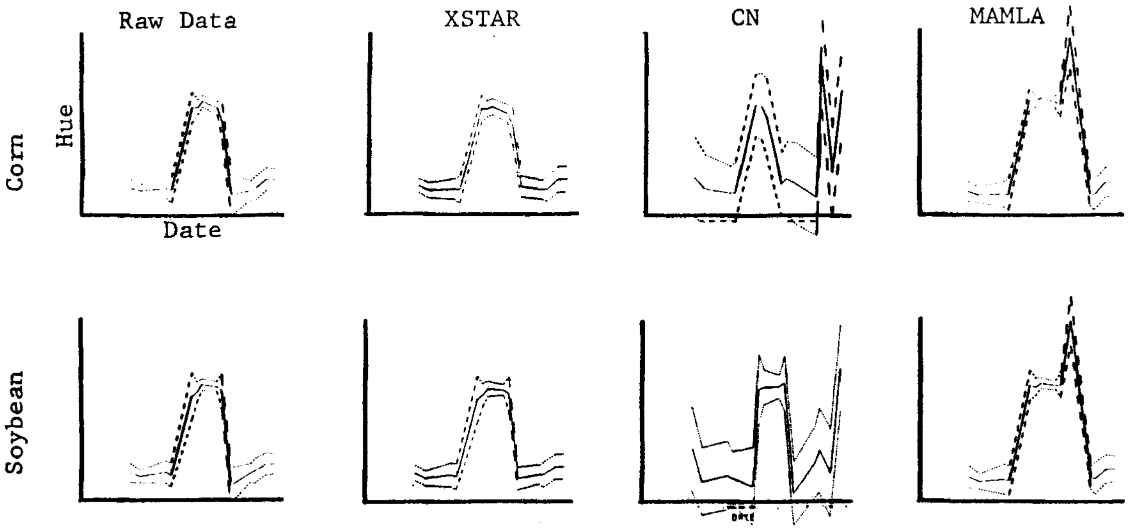


Figure 5. The Effects of Preprocessing on the Hue Profiles of Corn and Soybean.

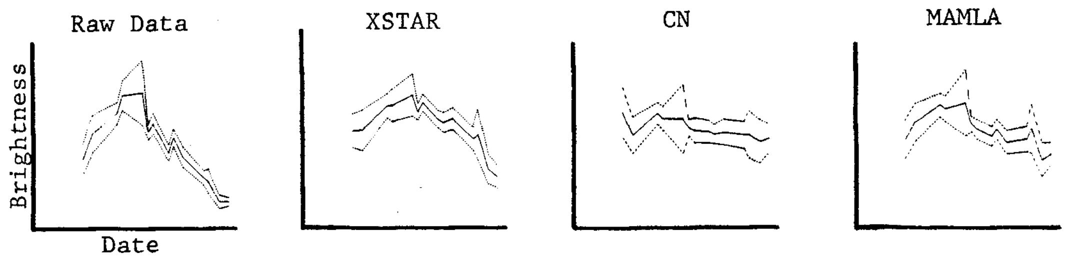


Figure 6. The Effects of Preprocessing on the Brightness Profiles of Trees.

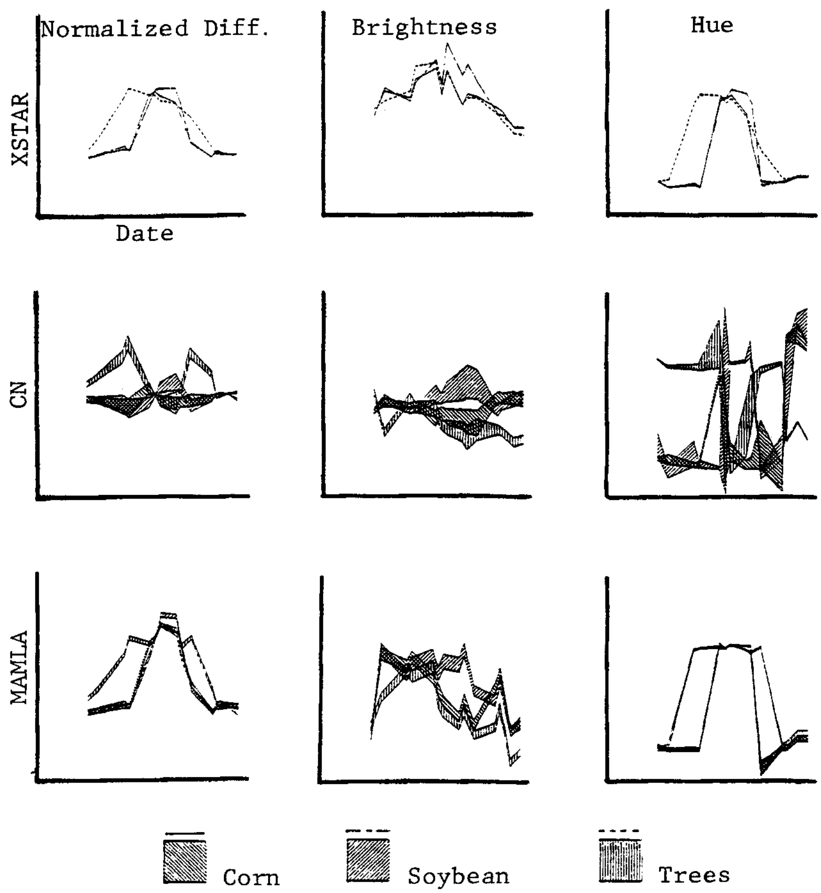


Figure 7. Sensitivity Envelopes of Normalized Difference, Brightness, and Hue.

Thomas M. Parris received a B.S. in Mathematics and Computer Science from the University of Michigan in 1982. He has been employed at ERIM since 1981, participating in the Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing (AgRISTARS) program under sponsorship of NASA's Earth Observation Division.

Richard C. Cicone's educational background is in the area of Mathematics and Computer Science having completed his M.S. at the University of Michigan in 1978. In nine years with the Environmental Research Institute of Michigan, his efforts have been predominantly in the area of civilian remote sensing. Working extensively for NASA JSC and GSFC, he has participated as researcher and manager in a variety of programs including SKYLAB, CITARS, NFAP, LACIE, AgRISTARS, and most recently LIDQA.