LARS Technical Report 071482

AN EVALUATION OF THEMATIC MAPPER SIMULATOR DATA FOR MAPPING FOREST COVER

M.E. DEAN, R.M. HOFFER

Purdue University/Laboratory for
Applications of Remote Sensing
West Lafayette, Indiana

ABSTRACT

This study* evaluated computer-aided analysis techniques applied to Thematic Mapper Simulator (TMS) data for the purpose of mapping forest cover types. Specifically, classification results obtained using a supervised set of training statistics and various combinations of three and four channels subsets of the seven available TMS channels are compared for three classification algorithms: L2, GML, and SECHO. In the analysis, the best three and four channel subsets were determined by minimum transformed divergence criteria. A Karhunen-Loéve or Principal Component linear transformation was applied to the 1979 TMS data set and supervised training statistics were generated for classifying the transformed data.

Classification results from applying the same three classification algorithms on the transformed data are compared to results from the untransformed data sets. Results from the untransformed TMS data show a higher performance using the four simulated Landsat channels (CH2:0.52 - 0.60μm; CH3: 0.63 - 0.69μm; CH4: 0.76 - 0.90μm; CH5: 1.00 -1 .30μm) than from the best four channels selected by the minimum transformed divergence criteria. The contextual classifier known as SECHO (Supervised Extraction and Classification of Homogeneous Objects) performed significantly better than either of the two per-point classifiers for the untransformed data. Overall classification results of the K-L transformation increased for the L2 algorithm, but decreased for both the GML and SECHO algorithms.

* This work was supported by NASA under Contract No. NAS9-15889.

I. INTRODUCTION

Extensive research and experience in the processing of MSS data for purposes in accurately classifying forest cover has been obtained in a wide variety of geographical regions. The new Thematic Mapper scanner system will have an increase in spectral and spatial resolution as well as an increase in the number of channels, which should theoretically allow better and more accurate classification of ground features, including forest cover types. Certain limitations may be encountered with this new system, however. Depending upon the particular scene characteristics, it is possible that higher interclass spectral variability may be introduced with the increase in spatial resolution, thus increasing the potential for spectral overlap and intraclass confusion. One problem with per-point classifiers, such as the GML (Gaussian Maximum Likelihood), commonly used in remote sensing applications, is that for spectral information alone, pixels within a particular cover class w_i may deviate from the class conditional pdf or probability density function $p(X|w_i)$, enough that they will be misclassified into another class. Preliminary work has shown that the increase in spatial resolution of the TM scanner can cause a decrease in performance over the current Landsat MSS system(5). Results from the use of contextual classifiers, such as SECHO (Supervised Extraction and Classification of Homogeneous Objects) which utilize both spectral and spatial association characteristics of the scene in the classification procedure, have indicated a potential for increasing TM classification performance(5).

In research dealing with computer classification of multispectral scanner data, consideration must be given to the trade-offs between classification accuracy and the cost of the analysis, such as with the computer time (CPU) required to ana-

lyze the data. For instance, classification time can be approximated by N(N+1) where N equals the number of features used in the classification sequence(12). It has also been shown that the cost of computer analysis (CPU time) increases disproportionately in relation to increases in classification accuracy beyond a certain optimum number of channels involved in the classification sequence(1). In addition, other studies have shown that an increase in the dimensionality of the feature space used in classifying MSS data will eventually result in a decrease in classification performance for a finite set of training statistics, due to the Hughes phenomenon(7).

It is obvious, therefore, that for scanner systems containing a large number of available wavebands (such as the Thematic Mapper), reduction of the feature space, while still retaining adequate classification performance, may be necessary in order to provide the user with an economical and therefore applicable resource management tool.

One such dimensionality reduction technique, the Karhunen-Loéve or Principal Component transformation, linearly transforms the sometimes highly correlated MSS data into an uncorrelated N-dimensional feature space oriented in such a way that the maximum data variance or information content is accounted for in descending order on the new transformed axes(9). Classification of the transformed data using the first two or three components is often comparable to results obtained when using more channels of the untransformed data(8).

II. OBJECTIVES

As indicated by the above discussion, there exists a potential for reducing the dimensionality of TM data through techniques such as the Principal Components Transformation or Feature Selection to define an appropriate subset of channels to use in the classification. However, it was not known how such reduction techniques would impact the performance of different classification algorithms. Therefore, the objectives of this study were defined as follows:

(1) To compare the effectiveness of two techniques (i.e., Feature Selection and Principal Components Transformation) that can be used to reduce the number of channels required for classifying Thematic Mapper Simulator data; and

Table 1. Descriptions of the various cover classes in the Camden test site.

Cover Class	Description
PINE	Pine forest areas, primarily plantations of slash and loblolly of varying age.
HDWD	Bottomland hardwoods such as sweetgum, willow, and bottomland oaks; mostly in dense old age stands.
TUPE	Water tupelo, primarily associated with narrow oxbow lakes and other areas of inundated soils.
CCUT	Areas subjected to clearcut forestry practices; clearcuts are in various stages of regrowth and may include windrowed slash.
PAST	Pastures and old fields.
CROP	Agricultural crops at various stages of development.
SOIL	Primarily areas of recently tilled agricultural fields, but may include some minespoil and recent clearcut areas.
WATER	Water areas include the Wateree River, small lakes and ponds, and turbid minespoil ponds.

(2) To compare the effectiveness of different classification algorithms, i.e.:

(i) L2 Minimum Euclidean Distance

(ii) GML (Gaussian Maximum Likelihood)

(iii) SECHO (Supervised Extraction and Classification of Homogeneous Objects

on both types of data sets (i.e., an original untransformed data set classified using the best three and four channel subsets determined by a common feature selection criterion and a data set transformed by a Principal Component Transformation using the first three and four components, respectively).

III. MATERIALS AND METHODS

A. DATA ACQUISITION

Data for this study consisted of aircraft multispectral scanner data obtained by NASA's NS001 Thematic Mapper Simulator (TMS). The wavelength bands on this scanner included three bands in the visible portion of the spectrum (CH1:0.45 - 0.52μm; CH2:0.52 - 0.60μm; CH3:0.63 - 0.69μm), two bands in the near IR (CH4:0.76 - 0.90μm; CH5:1.00 - 1.30μm), one band in the middle IR (CH6:1.55 - 1.75μm) and one band in the thermal IR

region (CH7:10.40 - 12.50μm). The data were obtained on May 2, 1979 over a study site in South Carolina near the city of Camden. The predominance of large contiguous tracts of forest (primarily bottomland hardwoods), in addition to minimal topographic relief made this a good site for this study. This area has also been designated by the U.S. Forest Service as one of two primary test sites for evaluating various remote sensing techniques for potential use in forest inventories. Table 1 provides a list of the designated cover classes in the Camden test site.

B. TRAINING AND TEST FIELD SELECTION

Training statistics were generated for the cover classes listed in Table 1 using a supervised approach(1). The "optimum" three and four channel subsets of the available seven channels were selected using a minimum transformed divergence criteria(12). Certain limitations associated with using such feature selection criteria include the fact that in the calculation of the transformed divergence, often class a priori probabilities are unknown and are therefore assumed to be equal, even though this is seldom the case. Further, there is no direct relationship between transformed divergence and the probability of error, although a lower bound can be determined for the divergence between two classes of equal a priori probability as follows:

$$\tfrac{1}{4}\exp(-\frac{D}{2}) \leq P_E \qquad (1)$$

Thus, it is possible that those classes with higher a priori probabilities may be discriminated against in favor of those classes of lower a priori probabilities and hence result in a lower overall classification performance.

Test fields of known cover types were selected through the use of a test grid of dimensions 50 lines by 50 columns. Test blocks, 25 by 25 pixels, were located in the southwest quadrant of each grid intersection and the largest possible field of every cover type present within that test block was selected and included in the test data set. By selecting test fields using this method it was assumed that the resulting test data set would be representative of the relative proportions of the various cover types present in the study area.

C. PRINCIPAL COMPONENTS

Due to the generally high degree of interband correlation between spectral bands of MSS data, the intrinsic dimensionality of the data, i.e., the dimensionality required to adequately describe the data, is often less than the original number of channels(9). One method for reducing the dimensionality of a particular data set by eliminating this interband correlation is to apply a common linear transformation known as the Karhunen-Loéve or Principal Component transformation to the data(6). The Karhunen-Loéve transformation calculates the eigenvectors associated with a sample covariance matrix of the data and thereby incorporates actual spectral variability inherent in the data in the transformation process. In essence, it rotates the sometimes highly correlated features in N dimensions to a more favorable orientation in the feature space, ordered such that the maximum amount of variance is accounted for in descending magnitude along the ordered components(6). Thus, the redundancy of information caused by correlation between bands is eliminated and a maximum amount of information content is concentrated onto a fewer number of axes. Figure 1 shows the information content associated with the various transformed components for the 1979 K-L transformed data set.

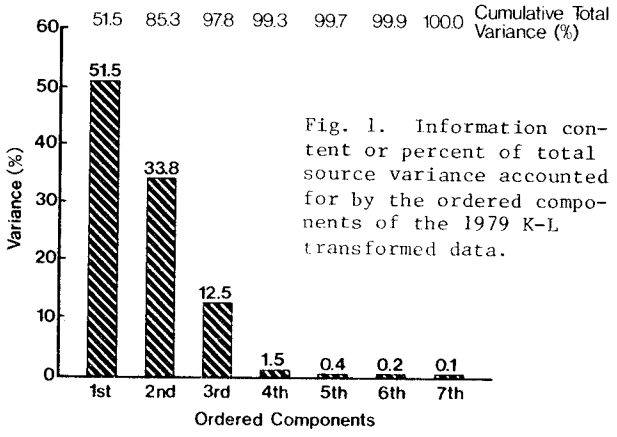

Fig. 1. Information content or percent of total source variance accounted for by the ordered components of the 1979 K-L transformed data.

The loadings or coefficients of the eigenvectors have been used in the past to describe the relative contributions of each original channel to the transformed channels and thus be used as another feature selection criterion. Caution must be observed using this approach, however, since this is primarily a heuristic approach which can only give indications as to the original value or contribution of each channel for a specific component.

Further, a high degree of interband covariance, and corresponding high correlation, may be reflected in the resulting coefficients of the two channels for a particular eigenvector; if both were to have relatively high coefficients, this may actually reflect their interband correlation rather than a significant and unique contribution from both.

Depending upon the eigenvalue associated with the ordered eigenvectors, i.e, the proportion of the total variance explained by a particular eigenvector and thus its overall importance, it may be possible to use two or three of the "significant" eigenvectors in order to determine the best two original channels. For instance, if two relatively uncorrelated channels were to both have relatively high corresponding coefficients for one of the first eigenvectors, i.e., the eigenvectors containing a significant amount of the total data source variance, then one could assume that they are each contributing a relatively significant amount of unique information. However, since the eigenvectors represent a linear combination of the original channel set, any uncoupling of these coefficients in order to determine their respective "contribution" to that eigenvector is heuristic and highly speculative.

D. FEATURE SELECTION

As mentioned in the previous section, feature selection techniques are primarily concerned with finding the optimum feature set which will adequately describe the intrinsic dimensionality of the data. Feature selection is of particular interest for purposes of minimizing computational time required to analyze data sets having significant dimensionality, i.e., large numbers of wavebands. Feature selection techniques for various pattern recognition applications have primarily been related to calculating bounds on the probability of error (and thus the probability of correct recognition = $1-P_E$(14).) Divergence as a measure of separability increases for decreasing P_E and lower bounds can be determined, as stated previously, although the direct relationship between the divergence and P_E is not well understood(12,14). Transformed divergence (TD) as a measure of probability of correct recognition tends to be a more ambiguous measure than other feature selection criterion, thus allowing a wide range of overlap in P_C (probability of correct classification) for a given TD value(14). Other less ambiguous measures of P_E include the Chernoff and Bhattacharyya bounds(14), although the computational complexities of these measurements restricts their practical use(3,13).

A minimum Transformed Divergence measure, TD(min), used in this study selected bands 1, 3, and 6 and bands 2, 4, 5, and 7 as the optimum three and four channel subsets, respectively.

E. CLASSIFICATION ALGORITHMS

The first classifier used in this study was the L2 or Minimum Euclidean Distance classifier. This is a relatively fast and therefore economical classifier which calculates the Euclidean or "straight-line" distance from a pixel to be classified to each of the mean vectors associated with the various cover classes, and then assigns the pixel to the "nearest" cover class:

$$\sum_{i=1}^{N} (X_i - M_{ij})^2 \qquad (2)$$

where: N = #channels

X_i = data value of pixel in channel i

M_{ij} = mean for class j in channel i

The L2 classifier does not take into account the spectral variation within each class and subsequently may not, depending upon the user's objectives, sufficiently minimize the probability of error.

The GML or Gaussian Maximum Likelihood algorithm is also a per-point classifier commonly used in remote sensing applications which calculates discriminant functions for each class from the associated class mean vectors and variance-covariance matrices. The GML algorithm is based upon the Bayes optimal strategy which produces results having the minimum probability of error over the entire data set for the given spectral information (12).

Decide $X \in w_i$ if and only if

$$g_i(X) \geq g_j(X) \text{ for all } i \neq j \qquad (3)$$

where: $g_i(X) = p(X|w_i)\ p(w_i)$

$p(X|w_i)$ = probability density function of X given X belongs to class w_i

$p(w_i)$ = a priori probability of class w_i

Due to the spectral variability of each cover class, there may be significant spectral overlap between the classes which could subsequently result in a relative high probability of error and misclassification.

The third algorithm, SECHO, is a contextual or per-field algorithm which first divides the scene to be classified into homogeneous fields and then classifies these fields using an extension of the GML algorithm(4). SECHO incorporates the fact that since cover classes are more likely to occur in homogeneous areas larger than one pixel in size (30m by 30m in this case), adjacent pixels are highly correlated, with the degree of correlation diminishing with an increasing distance between the pixels(4). Thus SECHO assigns an analyst-specified threshold value, below which adjacent pixels will be grouped into a homogeneous field. Statistics for these fields are calculated and compared to the original cover class statistics and a "homogeneous field" is classified as a unit into that class which it most closely resembles(4).

Table 2. Comparison of the overall classification performances between the untransformed TMS and K-L transformed data sets for all three classifiers.

Data Subset: "Best 3" Channels or 1st 3 Components

Classifier	Untransformed TMS[1] (Channels 1,3,6)	K-L Transformed Data (Components 1,2,3)
L2	65.2[a]	79.4[b]
GML	78.4[a]	82.4[b]
SECHO	86.8[a]	86.5[a]

Data Subset: "Best 4" Channels or 1st 4 Components

Classifier	Untransformed TMS (Channels 2,4,5,7)	K-L Transformed Data (Components 1,2,3,4)
L2	81.8[a]	84.8[b]
GML	88.1[b]	85.2[a]
SECHO	90.0[b]	87.8[a]

[1] Significantly different overall classification performances between data sets for each classifier is indicated by a different superscript (based upon a Newman-Keuls comparison with $\alpha = 0.10$).

III. RESULTS AND DISCUSSION

The K-L or Principal Components transformation was applied to the 1979 TMS data and then both the untransformed and transformed data sets were classified using both three and four channels (i.e., wavelength bands or transformed components) of data. The data were classified with the L2, GML, and SECHO classifiers, and in each case the results were evaluated using exactly the same test data set. The results are shown in Table 2. As this table shows, the results are mixed. The K-L transformed data using the first three components performed signficantly better overall than the untransformed TMS data set using the "best three" channels (1, 3, and 6), as selected by TD(min), for both the L2 and GML algorithms. With four components versus the "best four" channels (2, 4, 5, and 7), however, only the L2 performance increased significantly, while both the GML and SECHO algorithms decreased in performance with the transformation. Although not shown here, other three and four channel subsets of the original TMS data provided even better results than either the subsets of channels selected by TD(min) or the K-L components. In general, therefore, the K-L transformation provided a better three-channel feature set than that selected by TD(min), but not a better four-channel feature set in all cases. In addition, both the K-L transformation and TD(min) failed to provide the "optimum" three- and four-channel feature set for this particular data set.

The limitations of the TD separability as a feature selection criterion were previously discussed (i.e., the assumptions of equal class a priori probabilities and the degree of ambiguity associated with these measurements). One possible explanation for the failure of the K-L transformation to provide a distinct improvement in classification performance in all cases might be that some of the variance or information content of some of the less frequent cover classes, such as tupelo and crop, is being overwhelmed by the spectral variance or information content associated with the larger hardwood class. Since hardwood comprises the majority of the surface features in the test site, its input into the calculation of the transformation matrix was significant -- enough, perhaps, so as to "direct" the transformation in its favor and cause the spectral variability associated with the less frequent cover types to become reduced with the transformation. Certain algorithms such as the GML and SECHO may be more sensitive to this than

others. Therefore, in such cases, a K-L transformation may actually produce slightly worse results than with the original data. Other studies have shown the sensitivity of principal component analysis (PCA) for various cover features and how the more highly correlated the original data (which vary for different cover types), the higher the percentage of variance or information content that will be explained by a fewer number of components(2). Therefore, it might be better in certain cases for the analyst to define a supervised sample of data from which the transformation matrix can be calculated. This way, each of the features could be given the desired representation in the sample covariance matrix; the degree to which they would direct the transformation would then be related to their natural spectral variability.

Table 2 also shows that for both the untransformed and the transformed data sets, four channels (either wavelength bands or components) enable better classification accuracies to be achieved than three channels in every instance, provided the classifier is the same (i.e., for the L2 classifier, four channels result in higher classification accuracy than three channels, etc.). However, Table 2 also indicates that distinct differences between the classifiers were found. Therefore, a comparison was conducted to evaluate the statistical significance of these differences between the three classification algorithms, the results of which are shown in Table 3.

This table shows that in every case, except for four channels of transformed data, the GML performed significantly better than the L2 algorithm. In addition, in every case, SECHO performed better than either the GML or L2 per-point classifiers. However, detailed analysis of these results showed that these statements cannot be applied to all individual cover class performances; e.g., certain cover classes such as clearcut, crop and water; all performed better with the L2 classifier than for either the GML or SECHO algorithms. This may be due to relatively small variances in these three cover classes in comparison with the other cover types present; GML and SECHO would tend to classify pixels into those cover types with larger variances in order to reduce overall P_E (probability of error) even though the linear distance to the class means may be closer to a class of smaller variance. Further, all of the cover classes performed as well using the GML algorithm as with SECHO except for the hardwood category. Since hardwood com-

Table 3. Comparison of the overall and average class performances for three algorithms (L2, GML and SECHO) based on four data sets.

Data Set Description	Classification Performance (%) by Classifier					
	L2		GML		SECHO	
	Overall[1]	Average	Overall	Average	Overall	Average
3 Channels (1,3,6) Untransformed	65.2[a]	56.4	78.4[b]	70.4	86.8[c]	73.3
1st 3 Components, K-L Transformed	79.4[a]	74.4	82.4[b]	72.9	86.5[c]	75.1
4 Channels (2,4,5,7) Untransformed	81.8[a]	76.2	88.1[b]	78.5	90.0[c]	78.6
1st 4 Components, K-L Transformed	84.8[a]	71.6	85.2[a]	74.5	87.8[b]	73.4

[1] Different superscripts indicate significantly different overall classification performances between classifiers (based upon a Newman-Keuls comparison with $\alpha = 0.10$).

prises the greatest proportion of the test data, this was the primary reason for the greater overall performance of SECHO over GML. However, as the generally similar average class performances indicate, GML usually performed as well as SECHO, (the the major exception to this being the three-channel untransformed data). One distinct advantage of the SECHO classifier over the GML is the smaller amount of CPU time required to classify the data (as is shown in Table 4) and in the more interpretable classification maps obtained from SECHO; i.e., more uniform (homogeneous) fields of the various cover types are obtained than with the GML per-point algorithm.

Table 4. Classification time required for the L2, GML and SECHO algorithms to classify 10,000 pixels using four channels and 27 spectral classes.

	Classifier		
	L2	GML	SECHO
CPU (seconds)	28.9	82.6	51.6

Thus, from a practical standpoint, although the GML can perform as well as SECHO, the SECHO algorithm can provide an optimum trade-off between classification accuracy and cost of the analysis, as well as a more effective map output product.

IV. SUMMARY AND CONCLUSIONS

The results of this study can be summarized as follows:

* Transformed Divergence (TD) is an effective method for identifying various subset combinations of wavelength bands, thereby allowing the dimensionality of the data set used in the classification to be reduced, but the "Best n" subset should not be expected to always provide the optimum classification performance for both individual cover types and overall performance.

* Linear transformations, such as the Karhunen-Loéve (or Principal Component) transformation will condense the amount of data variability into a relatively small set of channels, as was shown in Figure 1. However, depending upon the relative proportions and spectral variability of the cover class samples which go into the calculation of the transformation matrix, the resulting separability of these classes in the transformed space may be less than in the original space and subsequently result in lower class and overall performances.

* A four-channel "optimum" subset of the total seven Thematic Mapper channels gave significantly better results than when using only three channels and, in general, enabled adequate class and overall performances to be achieved.

* Contextual classifiers such as SECHO, can obtain the same or better results than per-point classifiers such as the L2 and GML.

* The L2 classifier required the least amount of CPU time, with SECHO and the GML algorithms requiring sequentially greater amounts of CPU time for classification.

* The SECHO algorithm provided an optimum combination of classification performance, minimal CPU classification time, and output map product.

V. REFERENCES

1. Coggeshall, M.E. and R.M. Hoffer. 1973. Basic Forest Type Mapping Using Digitized Remote Sensor Data and ADP Techniques. LARS Information Note 030573, Laboratory for Applications of Remote Sensing (LARS), Purdue University, West Lafayette, IN 47906-1399. 131 pp.

2. Fontanel, A., C. Blanchet and C. Lallemand. 1975. Enhancement of Landsat Imagery by Combination of Multispectral Classification and Principal Component Analysis. Proc. NASA Earth Resources Resources Surv. Symp., July 1975, Houston, TX. NASA TMX-58168, pp. 991-1012.

3. Kailath, T. 1967. The Divergence and Bhattacharyya Distance Measures in Signal Selection. IEEE Trans. Comm. Tech., Vol. COM-15, No. 1, Feb. 1967, pp. 52-60.

4. Kettig, R.L. and D.A. Landgrebe. 1975. Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects. LARS Information Note 062375, LARS, Purdue University, West Lafayette, IN 47906-1399. 18 pp.

5. Latty, R.S. 1981. Computer-Based Forest Cover Classification Using Multispectral Scanner Data of Different Spatial Resolutions. LARS Technical Report 052081, LARS, Purdue University, West Lafayette, IN 47906-1399. 186 pp.

6. Morrison, D.F. 1967. Multivariate Statistical Methods. McGraw-Hill Book Co., NY. (2nd edition). 415 pp.

7. Muasher, M.J. and D.A. Landgrebe. 1981. Multistage Classification of Multispectral Earth Observational Data: The Design Approach. LARS Technical Report 101481, LARS, Purdue University, West Lafayette, IN 47906-1399. 170 pp.

8. Ready, P.J. and P.A. Wintz. 1972. Multispectral Data Compression Through Transform Coding and Block Quantization. LARS Information Note 050572, LARS, Purdue University, West Lafayette, IN 47906-1399. 148 pp.

9. Ready, P.J. and P.A. Wintz. 1973. Information Extraction, SNR Improvement, and Data Compression in Multispectral Imagery. IEEE Trans. Comm., Vol. COM-21, No. 10, Oct. 1973, pp. 1123-1130.

10. Schreier, H., L.C. Goodfellow and L.M. Lavkulich. 1982. The Use of Digital Multi-Date Landsat Imagery in Terrain Classification. Photogr. Eng. and Remote Sensing, Vol. 48(1), pp. 111-119.

11. Swain, P.H. 1980. A Quantitative Applications-Oriented Evaluation of Thematic Mapper Design Specifications. LARS Grant Report 121680, LARS, Purdue University, West Lafayette, IN 47906-1399. 16 pp.

12. Swain, P.H. and S.M. Davis (eds). 1978. Remote Sensing: The Quantitative Approach, McGraw-Hill Book Co., NY. 396 pp.

13. Swain, P.H., T.V. Robertson and A.G. Wacker. 1971. Comparison of the Divergence and B-Distance in Feature Selection. LARS Information Note 020871, LARS, Purdue University, West Lafayette, IN 47906-1399. 12 pp.

14. Whitsitt, S.J. and D.A. Landgrebe. 1977. Error Estimation and Separability Measures in Feature Selection for Multiclass Pattern Recognition. LARS Publication 082377, LARS, Purdue University, West Lafayette, IN 47906-1399. 186 pp.

AUTHOR BIOGRAPHICAL DATA

M. ELLEN DEAN is a research associate at the Laboratory for Applications in Remote Sensing (LARS), Purdue University, and is currently pursuing an M.S. in Remote Sensing through the Department of Forestry and Natural Resources at Purdue. She received a B.S. in Forestry from Virginia Polytechnic Institute and State University in 1979.

ROGER M. HOFFER is Professor of Forestry and Leader of Ecosystems Research Programs, LARS, Purdue University. B.S. in Forestry, Michigan State University; M.S. and Ph.D. in Watershed Management, Colorado State University. Co-founder of LARS in 1966, Dr. Hoffer currently teaches three courses in Remote Sensing of Natural Resources. He has lectured, consulted, and participated in remote sensing activities in many countries throughout North and South America, Asia, and Europe; has served as a principal investigator on Landsat, Skylab, and other major remote sensing projects, and has authored more than 130 scientific papers and publications on remote sensing. Professor Hoffer's research has focused on the use and refinement of computer-aided analysis techniques for forestry applications and on the study of the spectral characteristics of earth surface features.