paper to be presented at
49th annual meeting of
American Society of Photogrammetry
held Washington, D.C., March 1983

ANALYSIS OF LANDSAT FOR MONITORING VEGETABLES IN NEW YORK MUCKLANDS

M.H. Zhu, S.Y. Yan, W.R. Philipson, C.C. Yen and W.D. Philpot
Remote Sensing Program
Cornell University
Hollister Hall
Ithaca, New York 14853

BIOGRAPHICAL SKETCHES

Min-hui Zhu and Shou-yong Yan are Research Associates in the Chinese Academy of Sciences, People's Republic of China. Ms. Zhu is with the Institute of Electronics, Mr. Yan is with the Institute of Remote Sensing Application, and both are Visiting Scholars at Cornell University. Warren R. Philipson, Chain-chin Yen and William D. Philpot are with Cornell's Remote Sensing Program. Drs. Philipson and Philpot are on the faculty of the School of Civil and Environmental Engineering and Ms. Yen is a Data Analyst.

ABSTRACT

This pilot study assessed the feasibility of relying on Landsat multispectral scanner data for inventorying vegetables grown in mucklands, in variably shaped, variably sized fields. Classification of muckland vegetables using a Euclidean distance classifier and a parallelepiped classifier was performed with reasonable accuracy (generally over 60%) based on only one date of Landsat data. Prior canonical and principal component analyses did not improve the classification accuracy but did reduce the dimensionality of the data.

INTRODUCTION

Mucklands are important vegetable-growing areas in New York and other states. The area of muckland vegetables in New York is approximately 11,000 hectares (27,000 acres), which is nearly 40% of the total area of vegetables grown for fresh produce and 18% of the total area in vegetables.

The ultimate goal of this investigation is to develop an operational approach to monitoring the area of vegetables in New York mucklands. As envisioned, this would involve: (1) delineation of cultivated mucklands based on a priori information, (2) determination or development of the most effective algorithm for identifying different vegetables, and (3) development of the most effective procedure for measuring the area of the different vegetables each season.

The aim of the work described here was to address the second objective by determining the feasibility of distinguishing different vegetables with multispectral scanner data acquired by the Landsat satellites. Emphasis was placed on Landsat MSS data because they are acquired at regular or near-regular intervals, and because their basic cost would be lower than that of aircraft data for surveying mucklands scattered over the state. (As the cost of Landsat data rises, this may change.) Moreover, Landsat MSS data have been shown to be useful in crop surveys (e.g., Bizzell et al., 1975; Richardson et al., 1977; Hanuschak et al., 1979; NASA, 1979; Ryerson et al., 1979 and 1981; AgRISTARS Program Support Staff, 1982).

In general, most cropping studies have focused on large area, grain crops (Bizzell et al., 1975; NASA, 1979). Ryerson et al. (1979 and 1981) have dealt with the development of operational procedures for estimating the area of potatoes and beans in eastern Canada; however, they did not consider muckland vegetables.

MUCKLAND VEGETABLES

Found in virtually any climate, mucklands are characterized by shallow to deep, very poorly drained, organic soils, or "Histosols" (Soil Survey Staff, 1975). The soils are generally level or gently sloping, with a seasonally high water table. Permeability may be rapid in well-decomposed muck and variable in peaty materials. Prolonged wetness and rapid decomposition of the organic materials are the major limitations for farming. If adequately managed, however, muck soils are among the most productive, being well suited to intensive cultivation of row or specialized vegetable crops. The type of crop grown will be governed largely by climate.

A survey of muckland vegetables imposes severe constraints on satellite remote sensing systems. Water management of cultivated mucklands requires intensive surface and subsurface drainage. The grid of drainage canals delimits distinct field units, which are commonly small, narrow rectangular strips of variable orientation (Fig. 1). Field cropping management is such that one or many adjacent strips may be planted in a single crop. Although the individual field units are normally uniform in shape, at least one boundary of a block of units may be quite irregular if the boundary conforms with the natural boundary of the muckland.

METHODS AND MATERIALS

General Approach

An 11 July 1981 Landsat computer-compatible tape for central New York (scene #2236215030) was selected on the basis of a crop calendar and the available Landsat scenes. This tape was analyzed in batch mode on Cornell University's IBM 370/168 computer, using a series of multispectral data analysis routines developed at the Office of Remote Sensing of Earth Resources (ORSER), Pennsylvania State University. A 1977 version of the ORSER routines was used (Borden et al., 1977), the routines having been modified and supplemented over the years for operation on Cornell's system.

The Landsat data were related to field crop records supplied by the New York Crop Reporting Service. The pilot area encompassed 26 fields in Madison County, N.Y. (Table 1). The fields ranged in size from less than 1.0 to nearly 16.0 hectares. Field crop information was transferred to 1:24,000 scale U.S. Geological Survey topographic maps of the area (Canastota, Cleveland, Jewell and Manlius). Landsat brightness maps (ORSER "NMAPs"), geometrically corrected to a scale of 1:24,000, were generated for data matching. Attempts at unsupervised classification through a clustering algorithm produced poor results, and the analysis concentrated on supervised classification. Two types of supervised classifiers were used: "CLASS," a Euclidean distance classifier, and "PPD," a parallelepiped classifier.

Classification Routines

In ORSER's CLASS routine, classification of pixels is based on the nearest category mean. Each Landsat pixel is represented by four spectral values, which form a single vector or point in four-color space. The means of the four spectral values of the pixels which represent the different categories (e.g., onion or potato) in the scene are estimated on the basis of a priori or training set data; the latter being statistics derived from a sample of known pixels. Each category is thus represented by four mean spectral values or a centroid in four-color space. For classification, all pixels, including those used for training, are assigned to the nearest category centroid; they are classified as the centroid is classified if and only if they fall within a specified "critical distance" from the centroid. In ORSER, the critical distance is a *single* value for each category. It is arbitrarily defined, generally on the basis of the standard deviations about each category's mean spectral values.

In the present study, four categories of cultivated mucklands occurred in the study area: corn, onion, potato and abandoned fields. Four training areas were initially selected for each category. Classification was performed using the mean spectral values of the training areas for each category, and a critical distance calculated as the square root of the sum of the squared standard deviations about the mean spectral values. This was tried using several different combinations of training areas for the four categories. The best results were obtained by splitting the categories (Table 1); corn, onion and potato were each represented by two subcategories of two training areas each, and abandoned fields were represented by three subcategories of two training areas each (e.g., two training areas of "corn-1" and two of "corn-2").

Final implementation of the CLASS routine, which recognized the nine subcategories and a separate category for water, was performed in two steps. The first step used the mean spectral values and critical distances derived from standard deviations, as described previously. This approach classified all pixels into 11 classes: the nine subcategories, water and "other." The second and final step in classification relied on the results of the first step. The first classification output provided the Euclidean distances of separation among the centroids of the classified pixels. For the second classification, the critical distance for each class was recalculated as the average half distance between that class and all other classes.

Similar to other parallelepiped classifiers, implementation of ORSER's PPD is performed by defining the maximum and minimum spectral values for each category. Any pixel having values which fall within the parallelepiped defined for a certain category will be classified as that category, but only if the pixel values fall within only one category's parallelepiped. Pixels falling within more than one parallelepiped (i.e., overlapping parallelepipeds) will be classified as "confused." Parallelepipeds for the study area's nine subcategories and water were defined from the histograms of the spectral values of the training areas. These limits were refined until the best classification of training and testing area pixels was obtained.

In an effort to improve the accuracy of classification, pixels in the study area were subjected to canonical and principal component analyses prior to both CLASS and PPD classification. These analyses are described in a number of references (e.g., Morrison, 1976; Podwysocki et al., 1977; Jenson and Waltz, 1979).

RESULTS

The results of supervised classification using a Euclidean distance classifier, CLASS, and a parallelpiped classifier, PPD, with and without prior canonical or principal component transformations, are reported in Tables 2 through 5. The classifiers were applied to nontransformed data (labeled 0 in tables) and to transformed data from one to four canonical (Tables 2 and 4) or principal component (Tables 3 and 5) axes.

Reported in the tables is how the pixels representing each category (cover type) were classified--pixels used for training (upper half of tables) as well as known pixels set aside for testing (lower half). The percentages of training and testing pixels that were classified correctly are also reported along with the percentages of training and testing pixels that remained unclassified. The percentages of pixels that were classified incorrectly are the complements to these percentages and are not reported. In Tables 4 and 5, the percentages of pixels classified as "confused" by the parallelepiped classifier are also reported. This represents a special case of an incorrect classification, since a confused pixel is one classified as falling within the parallelepipeds of at least two categories, one of which might be correct.

DISCUSSION

The results are both promising and interesting. Using a single date of Landsat MSS data, it was possible to classify nontransformed pixels of muckland categories with accuracies ranging from 64% to 100% for training data and ~~54%~~ to ~~89%~~ for testing data. Although prior canonical or principal component analyses produced little or no increase in accuracy overall, these transformations did improve the classification of some categories and they did allow significant reduction in dimensionality, generally to two axes. 53-83%

Of particular interest is that classification accuracies with four axes of transformed data were substantially lower than classification accuracies with four spectral axes of nontransformed data. This outcome would not be possible if the classification parameters had not varied as the number of axes changed. Because of the specific limitations of the classification routines and the nature of the training data, however, the parameters were varied for each classification trial in an attempt to optimize the classification for each case.

As pixel values from the second, third and fourth canonical or principal component axis are added to those from the first axis for classification, the range in values may expand appreciably. Implementing the Euclidean distance classifier to operate on an increased number of axes of transformed data would thus require that the critical distance--a single value--be increased; otherwise an increasing number of pixels would remain unclassified. Increasing the critical distance to accommodate each added axis would likely decrease the resultant accuracy.

Accommodating additional axes would not affect the parallelepiped classifier in this manner, since the upper and lower limits for each dimension of a parallelepiped are selected independently. Adding a new dimension to a parallelepiped in order to incorporate a new axis in the classification would have no effect on the other dimensions. With the parallelepiped classifier, the decreased accuracies with additional axes are likely to have been caused by training areas of spectrally similar categories or training areas containing a number of incorrect pixels. In essence, the canonical transformation operates to separate pixels of different training areas; but if the different categories were spectrally similar or if some pixels were originally associated with the wrong category, error might eventually arise. Any error would be most pronounced in the third and fourth canonical axes. Although the principal component analysis operates on all data rather than just the training areas, the effect would have been the same. In this study, the same training areas were used for classification with both canonical and principal component analyses.

Although a more powerful classifier (e.g., maximum likelihood or an improved minimum distance classifier) might provide improved classification, the sensitivity of ORSER classifiers was observed, to an extent, serendipitously. Splitting the categories to achieve increased accuracy was performed empirically. Yet the subcategories offering the greatest improvement, corn and abandoned fields, were subsequently found to be physically different. With corn, the two subcategories represented fields having different densities of drainage canals; with abandoned fields, the three subcategories represented fields having drainage canals, no canals or trees.

CONCLUSIONS

Classification of muckland vegetable categories in New York could be performed with reasonable accuracy using only one date of Landsat MSS data. Canonical and principal component analyses did not improve the classification accuracy but did reduce the dimensionality of the data.

Future work with these and other muckland vegetable categories should consider temporal (multi-date) data and possibly other enhancements (e.g., ratios) and classifiers (e.g., maximum likelihood). Improved accuracy should also accompany the use of thematic mapper data from Landsat-4 (CORSPERS, 1976). This is assuredly true as regards the improvement in spatial resolution, though the value of the different spectral bands might first be confirmed through field study.

ACKNOWLEDGMENTS

This study was supported in part by NASA grant NGL 33-010-171 to Cornell University. Field cropping data were provided by Glenn W. Suter of the New York Crop Reporting Service, and the Landsat tape was provided by the Statistical Reporting Service, U.S. Department of Agriculture.

LITERATURE CITED

AgRISTARS Program Support Staff. 1982. AgRISTARS--Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing; Annual report, fiscal year 1981. Report AP-J2-04225. NASA, Houston, Tex. 59p.+appendices.

Bizzell, R.M., F.G. Hall, A.H. Feiveson, M.E. Bauer, B.J. Davis, W.A. Malila and D.P. Rice. 1975. Results from the Crop Identification Technology Assessment for Remote Sensing (CITARS) project. p.1189-1198. In Proc. 10th Int'l. Symp. Remote Sensing of Environ. Environ. Research Inst. of Mich., Ann Arbor, Mich.

Borden, F.Y., D.N. Applegate, B.J. Turner, B.F. Merembeck, E.G. Crenshaw, H.M. Lachowski and D.N. Thompson. 1977. Satellite and aircraft multispectral scanner digital data users manual. Tech. Report 1-77. ORSER, Penn. State Univ., University Park, Penn.

Committee on Remote Sensing Programs for Earth Resource Surveys (CORSPERS). 1976. Resource and environmental surveys from space with the thematic mapper in the 1980's. National Academy of Sciences, Washington, D.C. 122p.

Hanuschak, G. R. Sigman, M. Craig, M. Ozga, R. Luebbe, P. Cook, D. Kleweno and C. Miller. 1979. Obtaining timely crop area estimates using ground-gathered and Landsat data. Tech. Bul. 1609. U.S. Dept. Agr., ESCS, Washington, D.C. 29p.

Jenson, S.K. and F.A. Waltz. 1979. Principal components analysis and canonical analysis in remote sensing. p.337-348. In Proc. 45th Ann. Mtg. Amer. Soc. Photogram. A.S.P., Falls Church, Va.

Morrison, D.F. 1976. Multivariate statistical methods. 2nd ed. McGraw-Hill Inc., N.Y. 415p.

NASA. 1979. The LACIE symposium; A technical description of the Large Area Crop Inventory Experiment. (Held Oct. 1978) Report JSC-16015. NASA, Houston, Tex. 1125p.

Podwysocki, M.H., F.J. Gunther and H.W. Blodget. 1977. Discrimination of rock and soil types by digital analysis of Landsat data. NASA Preprint X-923-77-17. NASA/GSFC, Greenbelt, Md. 37p.

Richardson, A.J., C.L. Wiegand, R.J. Torline and M.R. Gautreaux. 1977. Landsat agricultural land-use survey. Photogram. Eng'g. & Remote Sensing 43:2:207-216.

Ryerson, R.A., R.J. Brown, J.L. Tamby, L.A. Murphy and B. McLaughlin. 1981. A timely and accurate potato acreage estimate from Landsat: Results of a demonstration. p.587-597. In Proc. 15th Int'l. Symp. Remote Sensing of Environ. Environ. Research Inst. of Mich., Ann Arbor, Mich.

Ryerson, R.A., P. Mosher. V.R. Wallen and N.E. Stewart. 1979. Three tests of agricultural remote sensing for crop inventory in eastern Canada: Results, problems and prospects. Canadian Jour. of Remote Sensing 5:1:53-66.

Soil Survey Staff. 1975. Soil taxonomy; A basic system of soil classification for making and interpreting soil surveys. Agr. Handbook 436. U.S. Dept. Agr., SCS, Washington, D.C. 754p. (see Chap. 11, "Histosols")

Figure 1. Portion of muckland vegetable study area in Madison County, New York

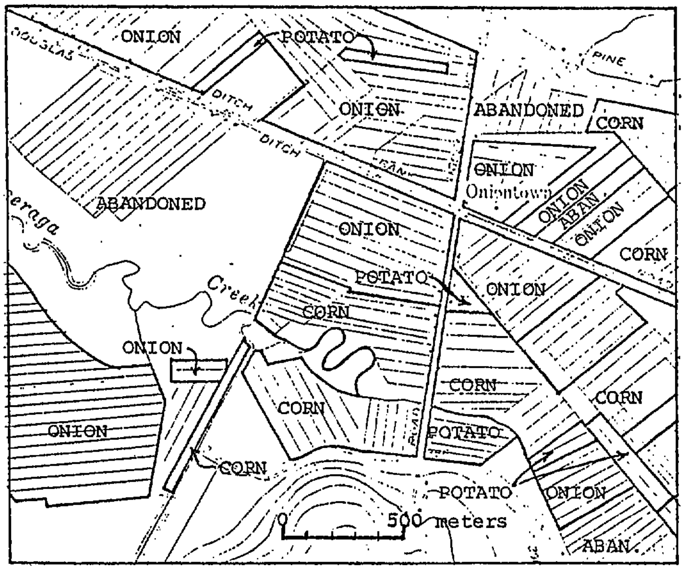

Table 1. Number of fields of each cultivated muckland used in classification.

COVER TYPE	NO. FIELDS IN TRAINING SETS	NO. FIELDS IN TESTING SETS	TOTAL FIELDS
Corn	4*	2	6
Onion	4*	2	6
Potato	4*	2	6
Abandoned	6**	2	8
			26

* Two subcategories of crop, two fields each.
** Three subcategories of cover type, two fields each.

Table 2. Supervised classification of pixels using a Euclidean distance classifier with and without prior canonical transformation.

	CLASSIFIED PIXELS																				
	CORN					ONION					POTATO					ABANDONED					
	NUMBER OF AXES*					NUMBER OF AXES					NUMBER OF AXES					NUMBER OF AXES					TOTAL
COVER TYPE	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	PIXELS
TRAINING SETS:																					
Corn	25	27	28	23	17	0	0	7	0	0	4	3	4	7	4	8	9	0	7	7	39
Onion	0	0	0	0	0	58	58	58	57	56	0	0	0	0	0	0	0	0	0	0	58
Potato	2	0	2	1	1	0	0	0	0	0	26	27	27	28	24	2	5	3	3	2	32
Abandoned	3	20	2	2	2	0	0	0	0	0	2	6	5	4	0	52	34	52	48	42	60
Correct (%)	64	69	72	60	44	100	100	100	98	97	81	84	84	88	75	87	57	87	80	70	
Unclass.(%)	5	0	0	5	28	0	0	0	2	3	6	0	0	0	16	5	0	2	10	27	
TESTING SETS:																					
Corn	18	6	18	13	11	6	7	5	6	0	5	4	7	7	6	0	17	3	3	1	34
Onion	0	3	3	2	4	29	30	29	28	28	4	0	0	0	0	0	0	0	0	0	33
Potato	0	1	4	3	2	0	0	0	0	0	19	18	23	20	14	7	13	6	7	6	33
Abandoned	2	2	2	1	1	0	0	0	0	0	11	11	16	17	8	25	30	24	20	11	43
Correct (%)	53	18	53	38	33	88	91	88	85	85	58	54	70	61	42	58	70	56	46	26	
Unclass.(%)	15	0	3	15	47	0	0	3	9	3	21	3	0	0	33	12	0	2	12	53	

* Denotes the number of canonical axes used in classification, with 0 being the classification of nontransformed data. Cumulative variation accounted for by the first through fourth axis was: 95.8%, 99.8%, 99.9% and 100%.

Table 3. Supervised classification of pixels using a Euclidean distance classifier with and without prior principal component transformation.

	CLASSIFIED PIXELS																				
	CORN					ONION					POTATO					ABANDONED					
	NUMBER OF AXES*					NUMBER OF AXES					NUMBER OF AXES					NUMBER OF AXES					TOTAL
COVER TYPE	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	PIXELS
TRAINING SETS:																					
Corn	25	27	20	20	18	0	0	0	0	0	4	3	6	5	6	8	9	13	11	9	39
Onion	0	0	0	0	0	58	58	58	57	55	0	0	0	0	0	0	0	0	0	0	58
Potato	2	0	2	1	1	0	0	0	0	0	26	26	28	26	25	2	6	1	1	2	32
Abandoned	3	22	4	3	2	0	0	0	0	0	2	6	3	1	0	52	32	53	49	45	60
Correct (%)	64	69	51	51	46	100	100	100	98	95	81	81	88	81	78	87	53	88	82	75	
Unclass.(%)	5	0	0	8	15	0	0	0	2	5	6	0	3	12	12	5	0	0	12	21	
TESTING SETS:																					
Corn	18	11	18	13	12	6	7	6	6	6	5	4	7	6	5	0	12	3	2	2	34
Onion	0	3	3	1	1	29	30	29	28	27	4	0	0	0	0	0	0	0	0	0	33
Potato	0	0	4	1	1	0	1	1	0	0	19	16	23	19	17	7	16	5	6	5	33
Abandoned	2	4	6	2	2	0	0	0	0	0	11	12	14	9	9	25	27	18	18	14	43
Correct (%)	53	32	53	38	35	88	91	88	85	82	58	48	70	58	52	58	43	42	42	33	
Unclass.(%)	15	0	12	20	26	0	0	3	12	15	21	0	0	21	30	12	0	12	33	42	

* Denotes the number of principal component axes used in classification, with 0 being the classification of <u>non</u>transformed data. Cumulative variation accounted for by the first through fourth axis was: 85.9%, 94.2%, 99.0% and 100%.

Table 4. Supervised classification of pixels using a parallelepiped classifier with and without prior canonical transformation.

	CLASSIFIED PIXELS																				
	CORN					ONION					POTATO					ABANDONED					
	NUMBER OF AXES*					NUMBER OF AXES					NUMBER OF AXES					NUMBER OF AXES					TOTAL
COVER TYPE	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	PIXELS
TRAINING SETS:																					
Corn	13	7	13	15	13	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	39
Onion	0	0	0	0	0	19	27	23	20	18	0	0	0	0	0	0	0	0	0	0	58
Potato	2	2	3	2	2	0	0	0	0	0	12	3	22	18	15	0	2	0	0	0	32
Abandoned	2	1	1	1	1	0	0	0	0	0	0	1	3	1	0	24	14	28	21	18	60
Correct (%)	33	18	33	38	33	33	47	40	34	31	38	9	69	56	47	40	23	47	35	30	
Unclass. (%)	38	8	28	49	51	55	12	40	52	62	25	0	9	31	41	53	3	37	57	3	
Confused (%)	26	74	36	13	10	12	41	21	14	7	31	78	12	6	9	3	70	10	5	65	
TESTING SETS:																					
Corn	10	4	12	7	7	2	2	2	2	3	0	1	1	0	2	0	0	0	0	0	34
Onion	0	0	0	0	0	19	20	19	17	12	0	0	0	0	0	0	0	0	0	0	33
Potato	2	5	2	2	3	0	0	0	0	0	10	3	11	8	8	1	0	3	4	2	33
Abandoned	2	1	2	1	1	0	0	0	0	0	6	2	8	8	5	8	14	8	8	7	43
Correct (%)	29	12	35	21	21	58	61	58	52	36	30	9	33	24	24	19	33	19	19	16	
Unclass. (%)	47	8	29	56	56	39	15	36	42	58	54	0	33	46	54	58	9	49	56	67	
Confused (%)	18	71	26	18	9	3	24	6	6	6	6	76	15	12	6	5	51	9	5	2	

* Denotes the number of canonical axes used in classification, with 0 being the classification of nontransformed data. Cumulative variation accounted for by the first through fourth axis was: 95.8%, 99.8%, 99.9% and 100%.

Table 5. Supervised classification of pixels using a parallelepiped classifier with and without prior principal component transformation.

	CLASSIFIED PIXELS																				
	CORN					ONION					POTATO					ABANDONED					
	NUMBER OF AXES*					NUMBER OF AXES					NUMBER OF AXES					NUMBER OF AXES					TOTAL
COVER TYPE	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	PIXELS
TRAINING SETS:																					
Corn	13	3	10	11	12	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	39
Onion	0	0	0	0	0	19	20	22	25	18	0	0	0	0	0	0	0	0	0	0	58
Potato	2	0	3	2	2	0	0	0	0	0	12	0	13	13	13	0	2	0	0	0	32
Abandoned	2	3	3	4	2	0	0	0	0	0	0	0	1	1	0	24	15	21	17	16	60
Correct (%)	33	8	26	28	31	33	34	38	43	31	38	0	41	41	41	40	25	35	28	27	
Unclass. (%)	38	15	28	46	54	55	14	36	48	66	25	0	16	22	34	53	2	35	50	62	
Confused (%)	26	77	46	26	15	12	52	26	9	3	31	94	34	31	19	3	68	18	13	8	
TESTING SETS:																					
Corn	10	1	14	12	11	2	1	1	3	3	0	0	0	0	0	0	0	0	0	0	34
Onion	0	0	0	0	0	19	19	19	15	13	0	0	0	0	0	0	0	0	0	0	33
Potato	2	1	7	3	3	0	0	0	0	0	10	0	11	13	11	1	1	1	1	1	33
Abandoned	2	0	5	3	2	0	0	0	0	0	6	0	7	3	2	8	15	8	9	10	43
Correct (%)	29	3	41	35	32	58	58	58	46	39	30	0	33	39	33	19	35	19	21	23	
Unclass. (%)	47	9	24	35	44	39	12	36	51	58	54	0	18	39	46	58	7	35	58	56	
Confused (%)	18	85	32	21	15	3	30	6	3	3	6	94	24	9	6	5	58	21	7	5	

* Denotes the number of principal component axes used in classification, with 0 being the classification of nontransformed data. Cumulative variation accounted for by the first through fourth axis was: 85.9%, 94.2%, 99.0% and 100%.