# CLASSIFICATION OF ECOLOGICAL AND CASI REFLECTANCE DATA FOR A FEN COMMUNITY IN NORTHERN MANITOBA

V. Thomas[1], P.Treitz[1,2], D. Jelinski[1,2], J. Miller[2,3], P.Lafleur[4], and H. McCaughey[1]

[1]Department of Geography, Queen's University, Kingston, Ontario, Canada K7L 3N6
Tel: (613) 533-6030; Fax: (613) 533-6122

[2]Centre for Research in Earth and Space Technology (CRESTech)
York University, 4850 Keele Street, Toronto, Ontario, Canada M3J 3K3

[3]Department of Physics and Astronomy, York University
4700 Keele Street, Toronto, Ontario, Canada M3J 1P3

[4]Department of Geography, Trent University, Peterborough, Ontario, Canada K9J 7B8

## ABSTRACT

Ordination and cluster analysis are two common methods used by ecologists to organize species abundance data into plant communities. When applied together they offer useful information about the relationships among species and the ecological processes occurring within a community. Remote sensing provides surrogate data for characterizing the spatial distribution of ecological classes based on the assumption of characteristic reflectance of species and associations. In this study, high spatial resolution CASI reflectance data were examined and compared to detailed ecological data for a fen environment in northern Manitoba. The goal of this research was to determine the relationship between ecologically-derived classes and reflectance. Ordination and cluster analysis techniques were used in conjunction with spectral separability measures to organize groups of community-based data that are suitable for classification of the CASI reflectance image. Results indicated that problems existed when attempting to relate TWINSPAN clusters to spectral reflectance. Higher classification accuracies were achieved when incorporating information from TWINSPAN, ordination, and signature separability analysis than by using any of these methods on their own.

## INTRODUCTION

Ecologists apply various methods (i.e. ordination and classification) for organizing ecological data into plant communities. The importance of ordination verses classification in ecology arises from the nature of the frequency and spatial distribution of species within communities. For instance, if community variation is discontinuous, and species organize themselves into discrete groups, then classification (clustering) is a natural framework for conceptualizing communities (Jongman *et al.*, 1995). On the other hand, if community variation is continuous, ordination is a more natural way of viewing these communities. Ordination and classification can be viewed as complementary, and when applied together offer useful information about the relationships among species and their distribution across sites.

This study focused on the integration of field collected species abundance data into the classification of high resolution CASI spatial-mode imagery. Of particular interest for this study was the multi-level hierarchical output of the commonly used TWINSPAN statistical package, and how the different levels varied in terms of spectral separability. TWINSPAN is a widely used statistical package amongst field ecologists, yet the relationship between species abundance/distribution across sites and spectral reflectance is not clearly understood. This is particularly important for imagery with very high spatial resolutions, where the potential exists for meaningful information to be derived from detailed ground information (Treitz *et al.*, 1992).

To assess the utility of field collected species abundance data for use with high spatial resolution CASI data, the following general questions were addressed.

- Can fen plant species abundance data be related to image spectral response?
- How useful is TWINSPAN clustering for this purpose? Do TWINSPAN divisions coincide with spectral separability of calibration sites?
- Does ordination provide useful information for grouping sites for image classification?

The purpose of this research therefore is to link the theories and practices of classification between the ecologist and remote sensing scientist.

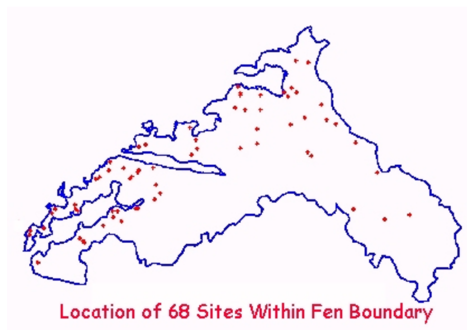
## STUDY AREA

The Northern Fen Site is one of several locations of interest in the Northern Study Area of the Boreal Ecosystems-Atmosphere Study (BOREAS), near Thompson, Manitoba. The fen is irregularly shaped and surrounded by water, with distinct patches of vegetation throughout.

## DATA DESCRIPTION

Field sampling throughout the fen catalogued over 114 different plant species at three canopy levels. This study utilized 68 visually distinct areas within the fen (each 3m x 3m), which were chosen and sampled during an intensive field campaign in 1995-96.

GPS coordinates were also collected for the center of each sample plot. The following diagram illustrates the general shape of the fen and the distribution of the 68 sites within it.




Species abundance data as percent cover were also collected for each plot, according to the following percent cover scheme.

| Level | % Cover |
|---|---|
| 1 | 0-1 |
| 2 | 1-5 |
| 3 | 5-25 |
| 4 | 25-50 |
| 5 | 50-75 |
| 6 | 75-95 |
| 7 | 95-100 |

**Table 1: Percent Cover Scheme**

In previous work, the sample plots were grouped into ecologically significant classes by using TWINSPAN analysis on the species abundance data. The first division of this analysis separated the sites into stands that were generally wet on the left side of the dendogram and generally dry on the right. Subsequent divisions separated the fen sites into the eight classes described below.

| Class | Characteristic Features |
|---|---|
| *Warnsturfia exannulatus / Carex rostrata* | Extremely wet sites found in the northern and western regions of the fen |
| Sedge Fen | Non-wooded, open stands composed predominantly of sedge. In some areas, this class is a transitional between *Warnsturfia exannlatus* lawns and shrub fens. |
| Shrub Fen | Characteristically dense shrub layer with no wooded canopy |
| Wooded Fen | Open canopy of stunted *L. laricina* present over a dense shrub layer |
| *Scorpidium scorpioides* carpets | Scorpidium scorpioes and Eriophorum alpinum are indicator species |
| Forested Fen | Occurs on drier sites with higher hummocks than wooded fens, although they are similar |
| Poor Fen Collapse | Associated with permafrost collapse scar |
| *Sphagnum fuscum* bogs | Moss layer typically covers greater than 75% |

**Table 2: TWINSPAN Groups**

CASI reflectance data were flown in July of 1996. These data were collected in spatial mode (0.5 m resolution), with six spectral channels approximately 10 nm wide, covering the visible and NIR portions of the electromagnetic spectrum (Table 3).

| Channel | Wavelength | Range |
|---|---|---|
| 1 | 463.63-474.05 | Blue |
| 2 | 525.07-535.55 | Green |
| 3 | 549.41-559.92 | Green |
| 4 | 630.90-650.48 | Red |
| 5 | 742.58-753.20 | NIR |
| 6 | 853.43-864.09 | NIR |

**Table 3: CASI Channel Wavelengths**

## PROCEDURES

### *Assessment of natural site arrangements*

To determine how the species and sites were naturally arranged, indirect gradient analysis was performed on the species abundance data. Species turnover across the sites (i.e. various sites had completely different species composition than one or more sites) suggested a unimodal response to latent environmental variables, which indicated that correspondence analysis (CA) was the most appropriate indirect ordination technique for these data (Jongman *et al.*, 1999; Cumming, 2000; Thioulouse *et al.*, 1997). Jongman *et al.* (1995) describe CA as a form of weighted averaging, which constructs a theoretical environmental variable that best explains the species data. This is done by maximizing the dispersion of species scores along the first ordination axis. This provides an easily interpreted bi-plot, where the natural grouping of sites would be indicated by their proximity to each other on the first ordination axis.

To explore the relationship between the natural species distribution shown in the CA output and image spectral response, canonical correspondence analysis (CCA) was performed to see how well spectral response described the natural species groupings. Where the CA described above maximizes dispersion of sites and species according to a latent environmental variable, CCA selects a linear combination of measured environmental variables to maximize the dispersion (Jongman *et al.*, 1995). The CCA used here incorporated all image channels (as environmental variables), as well as their standardized (correlation matrix) and nonstandardized (covariance matrix) principal components.

The natural arrangement of the fen sites was examined according to spectral response, by performing signature separability analysis comparing each site to all other sites. For this analysis, the Jeffries Matusita (J-M) distance algorithm shown below was used to calculate an easily interpreted measure of spectral separability between the sites (Schott, 1997; PCI, 1998).

(algorithm taken from PCI (1998) with modifications in terminology)

J-M(i,j) = 2*[1-exp(-a(i,j))]

J-M(i,j) = Jeffries Matusita distance between class i and j

A(i,j) = 0.125*T[M(i) –(j)]*Inv[A(i,j)*M(i)-M(j)+0.5*ln{det(A(i,j))/SQRT[det(S(i))*det (S(j))]}

J-M values ranging from zero to one indicated very poor separability between the sites (i.e. the sites had very similar spectral response). Values between 1.9 and 2.0 indicated that the two sites of interest had very different spectral response across the CASI bands (Schott, 1997; PCI, 1998).

### *Grouping the sites*

To group the sites based on the CA results, the ordination biplot was examined and the sites grouped based on separations along the first axis. This created groups that were similar in terms of species abundance and composition.

The J-M distance output was used to group the sites based on similarity in spectral response. By using the separability measure mentioned above for each site against all other sites, sites were grouped as follows.

- Sites found to be completely spectrally separable (i.e., J-M distance > 1.9) from each other were never grouped together.
- If a site was shown to be very poorly separable (i.e., J-M distance < 1.0) with different sites, all of those sites were grouped together. The class was then checked for conflicts by ensuring that no combination of sites within the class had good separability.

Use of the above guidelines enabled the grouping of 27 of the original 68 sites into 5 initial classes. The remaining 41 sites were shown in the separability analysis to be moderately separable. These sites were then added to the class from which they were most poorly separable. Groupings were based on the following.

- If a site had poor separability from any of the original groups, it was added to that group, provided there were no conflicts with any member of that group.
- If multiple sites had good separability from all of the original groups, a new class was created. This resulted in 5 new classes.
- Sites that had good separability with all or most classes were not grouped (one site).
- Sites which could not be placed into any group without conflict were not grouped (four sites).

### *Assessment of original groups*

Groups created using CA and J-M distance analyses were used to assess the suitability of TWINSPAN image classification of the fen. To accomplish this, each level of the dendogram was examined, starting at the arbitrary first hierarchical break. This was first done by examining the CA biplot to determine where on the first CA axis the first TWINSPAN division occurred.

One of the potential problems with TWINSPAN is that sites close to this division, but on opposite sides, will never be grouped together regardless of how ecologically similar they may be (Jongman *et al.*, 1995; Cumming, 2000; Hill, 1979; van Groenewoud, 1992). Sites that fell near the first division on the CA axis were identified as potential problems and examined to determine if they were spectrally separable. In terms of image classification, there would only be a significant problem if sites were divided that were not spectrally separable.

Problems that occurred at the first (i.e., the highest) division were carried on throughout all of the lower divisions. To determine the extent of the potential problems, each level of the hierarchy was examined to determine where in the TWINSPAN divisions conflicts with spectral separability existed.

### *Supervised image classifications, error analysis, and derivation of new groups*

Once initial groups were created with the CA and J-M Distance analysis, a variety of maximum likelihood classifications were performed in the Environment for Visualizing Images (ENVI). As discussed, the main objective of this analysis was to incorporate the TWINSPAN and/or CA groupings in the image classification, in hopes of achieving a higher classification accuracy than that which was achieved by any of these means alone. Overall classification accuracy was the preferred measure to assess this (rather than average accuracy), since it weights the accuracy of each class by the proportion of test samples for that class in the training or validation set (Research Systems, Inc., 1999; PCI, 1998). As well, Kappa statistics were generated to describe the proportion of agreement between the classification result and the validation sites after random agreements by chance are removed from consideration (Rosenfield and Fitzpatrick-Lins, 1986; Richards, 1986)

Once the classification accuracies for the individual methods were determined, the results were examined in detail to create more optimal classes. This involved the examination of the commission and omission errors in the confusion matrixes to determine the amount of validation pixels that were incorrectly classified. Groupings from the best classifications that achieved high accuracy and were related to meaningful ecological groupings were left unchanged. Individual sites that were identified through J-M distance analysis to be potential problems were analyzed separately. If a class contained two or more TWINSPAN groups, the class was subdivided to reflect this. In this manner, new classes were created containing groups of successfully classified sites (that had ecological meaning) or individual sites whose grouping was unclear.

Once a first derivation of optimal classes was achieved, separability analysis was performed on each class. Again, classes that were not spectrally separable were merged, while those that were very well separated were left. This logic was applied in an iterative fashion to group all sites (including extreme sites) in a spectrally separable but ecologically meaningful way.

## RESULTS & IMPLICATIONS

Due to the iterative nature of the techniques utilized in this analysis, the results are too numerous to be reported here. Rather, comments will focus on the final optimal groupings and observations made throughout the process to derive these groups.

The final classification based on groupings derived from TWINSPAN, CA, and signature separability analysis resulted in 8 groups which were related to the original TWINSPAN classes as shown in Table 4.

| Group | Ecological Description |
|---|---|
| 1 | Dominantly forested. All sites on the right side of the first TWINSPAN division. |
| 2 | Combination of forested and wooded communities. Sites located on both sides of 1st TWINSPAN division. Could not be derived from TWINSPAN, but sites are located closely together on the first CA axis. |
| 3 | Combination of sedges and shrubs. |
| 4 | Dominantly *Warnstorfia exannulatus/ Carex rostrata* lawns. |
| 5 | Combination of poor fen collapse, *Sphagnum fuscom* bogs, and *Scorpidium scorpioides* carpets. Not creatable with TWINSPAN or CA. Both sides of the first TWINSPAN division. |
| 6 | Combination of sedges, wooded communities, and *Scorpidium scorpiodes* carpets. Left side of the first TWINSPAN division. |
| 7 | Combination of *Warnstorfia exannulatus/ Carex rostrata* lawns and sedges. Left side of the first TWINSPAN division. |
| 8 | Combination of *Warnstorfia exannulatus/ Carex rostrata* lawns, sedges, shrubs, and wooded communities. Left side of the first TWINSPAN division. |

**Table 1: Ecological Description of Final Groups**

It is evident when examining the Table 4 that, in terms of image classification, there were problems with the TWINSPAN groupings even at the first division. This was particularly noticeable in the second group above, which indicated a similar spectral response for various sites on both sides of the TWINSPAN diagram. The implication of this finding is that TWINSPAN could never group these sites in an optimal way for image classification. However, it was interesting to note that the CA biplots did illustrate that these sites were actually very similar in terms of species abundance and distribution.

The best final overall classification accuracy for the optimally derived groups was 62.4%, with a Kappa coefficient of 0.55. This value was 22% higher than the best TWINSPAN classification when performed in ENVI.

Note that the relatively low Kappa statistic was consistent with the findings of the CCA analysis, which indicated that the image channels and principal components of these channels did not relate to the species distribution across sites very well. The key results in this analysis revealed an explained variance of only 44%, with $r^2$ =0.47.

Producer's accuracy for individual classes in the best final result ranged from 33% to 93.5%. User's Accuracy for individual classes in the best final result ranged from 39% to 72.3%.

Note the relatively wide range of classification accuracies across the various classes. The highest class accuracy was for a group composed of a combination of sedges and shrubs. It was not surprising that there would be some spectral similarities between these two ecological groupings, due to the similarities in species composition between these groups.

It was also interesting to note that the lowest class accuracy was for a group of sites composed of sedges, wooded communities and *Scorpidium scorpioides* carpets. Although the species composition of these three groups was quite different, the spectral separability analysis indicated that these sites were spectrally similar. This case also emphasized the importance of the selection of calibration vs. validation sites for the classification accuracy assessment. One would expect a source of error if the various ecological groups were not represented evenly in both the training and validation data sets.

Based on the results from several stages in the analysis, it could also be seen that:

1. Classification results from TWINSPAN groups alone were higher than results from groups that were interpreted from CA alone. Classification performed on the CA groups gave the lowest accuracy of all grouping methods.
2. Extreme sites (sites with very different species composition) were very problematic for grouping purposes. Both TWINSPAN and CA results indicated that the extreme sites could be classified together. However, J-M distance analysis illustrated that these sites should not be in the same spectral group, but that it was possible (but not optimal) to fit the extreme sites into other groups.

## SUMMARY

Based on the results discussed above, the following general summary statements can be made regarding the relationship between species abundance, distribution, and spectral characteristics for high spatial resolution CASI imagery.

1. TWINSPAN groups were not optimal for classification of this fen. At all stages in the analysis, there were site that were separated by a TWINSPAN division that were not spectrally separable.
2. Most optimal TWINSPAN results occurred at multiple levels in the TWINSPAN divisions.
3. CA proved to be very useful when examined in conjunction with J-M groups and the TWINSPAN output. It identified problem separations in the TWINSPAN output and helped to subdivide TWINSPAN groups when the separability analysis indicated a problem within a branch of the dendogram.

Finally, it was shown that using CA bi-plots, TWINSPAN clustering, and J-M analysis together provided a higher classification accuracy than could be achieved by any one method alone. This suggests that incorporation of species abundance data for ecological clustering/ordination can assist in spectral classifications of fens when using high spatial resolution data.

## REFERENCES


Cumming, B. 2000. Department of Biology, Queen's University. Personal Communication.

Glaser, P.H. 1989. "Detecting biotic and hydrogeochemical processes in large peat basins with Landsat TM imagery," *Remote Sensing of the Environment*, Vol. 28, pp. 109-119.

Hill, M.O. 1979. *TWINSPAN. A fortran program for arranging multivariate data in an ordered two-way table by classification of the individuals and their attributes.* Microcomputer Power. New York.

Jongman, R.H.G., Ter Braak, C.J.F. and O.F.R. van Tongeren. 1995. *Data Analysis in Community and Landscape Ecology*. Cambridge University Press, Cambridge, U.K.

PCI. 1998. *ImageWorks Version 6.3 EASI/PACE.* PCI Geomatics, Richmond Hill, Canada.

Research Systems, Inc. 1999. *ENVI User's Guide*. Better Solutions Consulting Limited Liability Company, Boulder, Colorado.

Richards, J.A. 1986. *Remote Sensing Digital Image Analysis.* Springer-Verlag. New York.

Rosenfield, G.H., and Fitzpartick-Lins, K. 1986. "A coefficient of agreement as a measure of thematic classification accuracy," *Photogrammatics Engineering of Remote Sensing*, Vol. 52, No. 2, pp. 223-227.

Schott, J.R. 1997. *Remote Sensing The Image Chain Approach*, Oxford University Press. Toronto.

Thioulouse, J., D. Chessel, S. Doledec, and J.M. Olivier, 1997. "ADE-4: a multivariate analysis and graphical display software," *Statistics and Computing*, Vol. 7, pp. 75-83.

Treitz, P.M., Howarth, P.J. and R.C. Suffling. 1992. "Application of Detailed Ground Information to Vegetation Mapping with High Spatial Resolution Digital Imagery," *Remote Sensing of the Environment*, Vol. 42, pp. 65-82.

van Groenewoud, H. 1992. "The robustness of correspondence, detrended correspondence, and TWINSPAN analysis," *Journal of Vegetation Science*, Vol. 3, pp. 239-246.