

Image classification of a northern peatland complex using spectral and plant community data

Valerie Thomas^{a,*}, Paul Treitz^a, Dennis Jelinski^{a,b}, John Miller^c,
Peter Lafleur^d, J. Harry McCaughey^a

^aDepartment of Geography, Queen's University, Kingston, Ontario, Canada K7L 3N6

^bSchool of Environmental Studies, Queen's University, Kingston, Ontario, Canada K7L 3N6

^cDepartment of Physics, York University, 4850 Keele Street, Toronto, Ontario, Canada M3J 1P3

^dDepartment of Geography, Trent University, Peterborough, Ontario, Canada

Received 12 March 2001; received in revised form 5 June 2002; accepted 5 June 2002

Abstract

Ordination and cluster analysis are two common methods used by plant ecologists to organize species abundance data into discrete “associations”. 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 species associations. Currently, there exists a need to establish and clarify the link between theories and practices of classification by ecologists and remote sensing scientists. In this study, high spatial resolution Compact Airborne Spectrographic Imager (CASI) reflectance data were examined and compared to plant community data for a peatland complex in northern Manitoba, Canada. The goal of this research was to explore the relationship between classification of species cover and community data and reflectance values. Ordination and cluster analysis techniques were used in conjunction with spectral separability measures to organize clusters of community-based data that were suitable for classification of CASI reflectance data, while still maintaining their ecological significance. Results demonstrated that two-way indicator species analysis (TWINSPAN) clusters did not correspond well to spectral reflectance and gave the lowest classification results of the methods investigated. The highest classification accuracies were achieved with ecological classes defined by combining the information obtained from a suite of analysis techniques (i.e., TWINSPAN, correspondence analysis (CA), and signature separability analysis), albeit not statistically superior to the classification obtained from the signature separability analysis alone.

© 2002 Elsevier Science Inc. All rights reserved.

1. Introduction

Boreal forest peatland is a major circumpolar ecosystem (Elliot-Fisk, 2000) comprising about 15–20% of Canada's boreal forest landscape (NWWG, 1988). The importance of boreal forest peatland ecosystems has been recognized by scientists from many disciplines (e.g., Calmé & Desrochers, 2000; Lavoie & Saint-Louis, 1999; Mazerolle, Drolet, & Desrochers, 2001; Waddington & Warner, 2001). For instance, peatlands provide a historical record essential to research on the climate, biology, and culture of the areas in

which they exist (Gorham, 1991; Lafleur, McCaughey, Joiner, Bartlett, & Jelinski, 1997; Lavoie, Zimmermann, & Pellerin, 2001; Petrone, Waddington, & Price, 2001).

The two major types of vegetation present in peatlands are ombrotrophic bog and minerotrophic fen plant communities. The separation of bogs and fens is based on (1) underlying landform; (2) water chemistry; (3) hydrology; and (4) indicator species. Bogs are often raised in elevation, low in pH, poor in species diversity, and low in calcium (Ca). In contrast, fens tend to have concave landforms, are less acidic, have richer floral species assemblages, and are higher in Ca (Glaser, 1992; Mitch & Gosselink, 1993; Wright, Coffin, & Aesang, 1992). These differences are linked to the hydrological properties of fens and bogs. Fens are further subdivided into poor fen, intermediate rich fen,

* Corresponding author. Tel.: +1-613-533-6030; fax: +1-613-533-6122.

E-mail address: valthomas@yahoo.com (V. Thomas).

and extremely rich fen, whereby pH, Ca concentration, and species diversity increase along a gradient. It is not unusual within large fens to find all three fen subtypes owing to variations of within-fen hydrological and nutrient regimes.

Fens and bogs sequester carbon via peat accumulation. This is accomplished through primary productivity from the uptake of atmospheric carbon dioxide (CO_2). In some cases, these ecosystems are also important as emitters of CO_2 , as well as methane (CH_4), which makes them an important concern for climate change (Roulet, 2000; Schlesinger, 1997). Studies of carbon cycling and trace gas biogeochemistry have shown that plant community composition and wetland microtopography (hummocks and hollows) are important controls on the exchange of these trace gases (Bubier, Crill, Moore, Savage, & Varner, 1998; Bubier, Moore, & Roulet, 1993; Moore, Heyes, & Roulet, 1994; Waddington & Roulet, 1996; Whiting, 1994). Thus, heterogeneity and patchiness of vegetation within peatlands has important consequences for CO_2 and CH_4 fluxes and creates a large range in trace gas fluxes that can ultimately be linked to thermal conditions and local scale hydrology (Potter, Bubier, Crill, & Lafleur, 2001). For example, CH_4 fluxes vary from hummock to carpet and pool, corresponding with a decrease in height above the mean water table position (Bubier, Moore, & Juggins, 1995); similar results are found for CO_2 fluxes (Waddington & Roulet, 1996). The close correspondence between peatland topographic position and species assemblage may be used as a surrogate for the degree of anaerobism/aerobism in a peatland, and hence for use in estimating CO_2 and CH_4 fluxes. Further, previous studies suggest that it may be possible to identify these community differences from remote sensing signatures (Bubier, Rock, & Crill, 1997; Whiting, 1994), providing a remote sensing tool for estimating peatland gas exchange.

The Boreal Ecosystem–Atmosphere Study (BOREAS) (Sellers et al., 1995) was an intensive remote sensing and field study of the northern boreal forests of Canada. BOREAS investigated exchanges of energy, water, heat, CO_2 , and trace gases between the boreal forest and the atmosphere. Remote sensing science was used to develop linkages between spectral response and boreal zone biophysical processes that govern these exchanges at a range of spatial scales. Here, we report on the application of Compact Airborne Spectrographic Imager (CASI) data to generate reflectance images of fen and bog vegetation for comparison to field surveys of species composition.

Plant community ecologists often analyze vegetation data by a methodological duet consisting of ordination and classification (as well as direct gradient analysis) (Gauch, 1982). Both ordination and classification techniques organize community data on species abundances independent of the habitat template. The choice of ordination versus classification for any particular plant data set is not straightforward. In general, if community variation is discontinuous, and species organize themselves into discrete clusters, then classification (clustering) is a natural framework for con-

ceptualizing communities (Jongman, Ter Braak, & van Tongeren, 1995). However, data structures that are naturally strongly clustered are rare in community ecology (Gauch, 1982). Moreover, even when a data set and ecological environment appear to be good candidates for classification, it remains that classification is partly an art, whereby the investigator's experience and understanding are of paramount importance (Gauch, 1982). On the other hand, if community variation was continuous, ordination would appear to be a more logical method for analysis of community structure. However, special problems arise with ordination in that it is descriptive in nature and subject to a number of assumptions about joint relationships of variables. Ordination and classification may be viewed as complementary, and when applied together, offer useful information about the relationships among species and their distribution across sites.

Of particular interest in this study is the application of two-way indicator species analysis (TWINSPAN) (Gauch & Whittaker, 1981; Hill, 1979) to divide samples of species into a hierarchy of statistically similar clusters and then to examine their spectral separability. TWINSPAN is widely employed by field ecologists, yet the relationship between species abundance/distribution across sites and spectral reflectance is not clearly understood. Good correspondence between clusters of ground sampled vegetation and spectral signatures at fine spatial scales would imply that high resolution spectral data could be used in the generation of region-wide model estimates of trace gas flux, especially when up-scaled to coarser resolution spectral data.

Ordination analysis is also employed in this study, such that species and samples are arranged in a low-dimensional space whereby similar species are nearby and dissimilar entities far apart. Here, the degree to which separability is improved by use of correspondence analysis (CA) is analyzed (Ludwig & Reynolds, 1988). Correspondence analysis ascertains the degree of ecological “correspondence” between sampling units and species using an eigen-analysis approach.

Another method of determining the natural arrangement of the plant communities within a peatland is through the examination of their spectral similarities and differences using spectral separability analyses (Anderson & Clements, 2000). Separability measures (e.g., divergence, transformed divergence, Jeffries–Matusita (J–M) distance) are common in the analysis of multispectral remote sensing data (Jensen, 1996; Mausel, Kamber, & Lee, 1990; Swain & Davis, 1978; Treitz & Howarth, 2000). Here, Jeffries–Matusita (J–M) distance is applied to remote sensing spectral data to assess plant communities within the peatland complex.

The overarching objective is to determine which of these methods used separately, or in combination with one another, can provide the most accurate characterization of community-level structure in peatlands based on their spectral properties. This study integrates species abundance data into the classification of high spatial resolution CASI data.

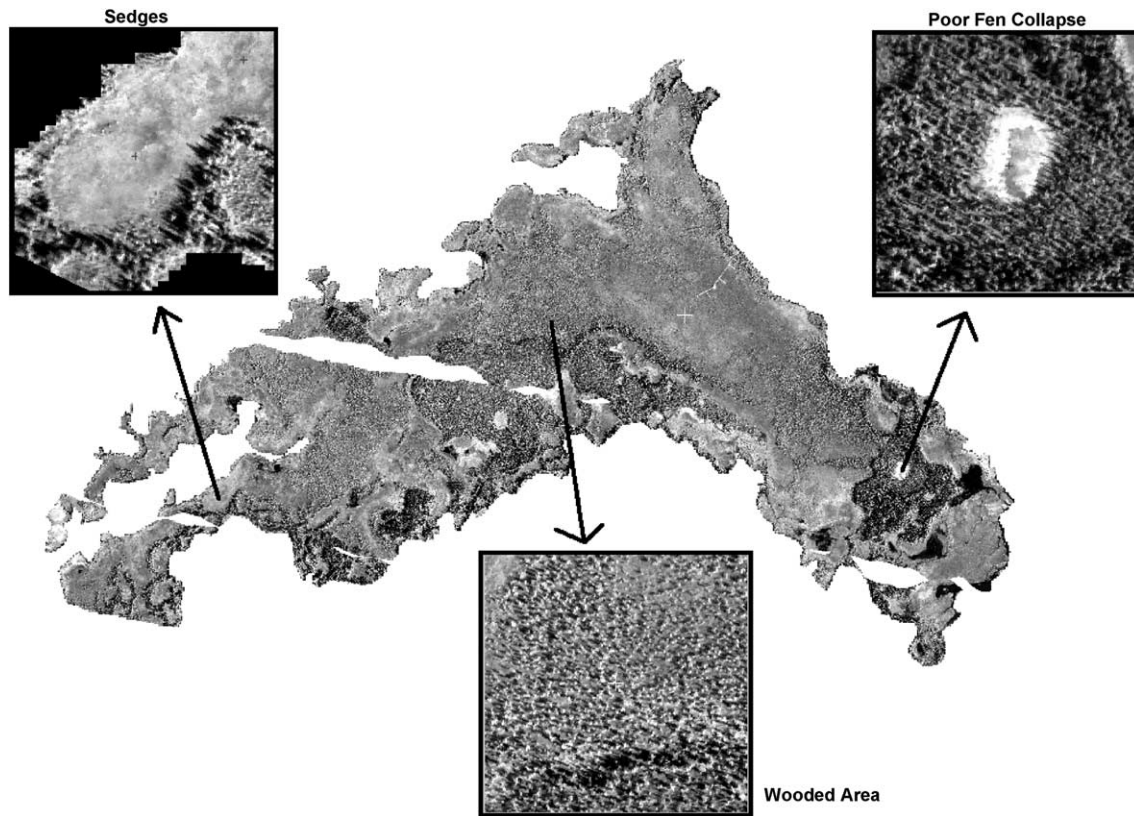


Fig. 1. Gray-scale of the peatland complex, with community subsets. The peatland is one of several sites at the BOREAS northern study site. This site is located at 55.9°N 98.4°W.

This combination of techniques is important for analyzing imagery with very high spatial resolutions, where the potential exists for meaningful information to be derived from detailed ground information (Anderson & Clements, 2000; Jacobsen, Nielson, Ejmaes, & Groom, 1999; Treitz, Howarth, & Suffling, 1992). Therefore, the overall goal of this research is to explore the link between the theories and practices of classification of vegetation data by ecologists and image classification for mapping the spatial extent of vegetation by remote sensing scientists.

2. Study area

The peatland site is one of several locations of interest in the BOREAS northern study area near Thompson, Manitoba, Canada (55.9°N, 98.4°W). Details of the site's micro-meteorology, general vegetation, and hydrology are described by Lafleur et al. (1997). Briefly, the peatland has an area of approximately 500 ha and is kidney-shaped. It is underlain by 1–6 m of peat, with peat depths greater than 5 m in the central portion (Jelinski, unpublished data). In general, the peatland platform is heterogeneous in vegetation character with an undercover of bryophytes (mosses), a subcanopy of bog birch (*Betula glandulosa*), a range of sedges species (*Carex* spp.), and sparsely scattered tamarack

(*Larix laricina*). At a finer scale, the vegetation is characterized by large patches of visually distinct vegetation, interlaced with less discrete transitional ecotones (Fig. 1). These patches can be broadly categorized into four community types. The wettest community is treeless and dominated by sedges and buck bean (*Menyanthes trifoliata*). *Drepanocladus exannalatus*, a brown bryophyte, is the dominant understory species. The sedge-fen community type is somewhat drier and is dominated by *Carex* spp. and *M. trifoliata*. The moss-shrub community is drier, hummocky, and dominated by bog birch and bog rosemary (*Andromeda porifolia*). Small stunted tamaracks are scattered in the understory. The driest site is characterized by a relatively dense overstory of tamarack and a thick understory carpet of bryophytes such as *Sphagnum warnstorffii*, *S. angustifolium*, and *Tomenthypnum nitens*. Standing dead or decadent tamarack are common in parts of the peatland.

3. Methods

3.1. Multivariate analysis techniques

3.1.1. Two-way indicator species analysis (TWINSPAN)

TWINSPAN has been implemented as a computer program that divides the original data set into smaller subsets

according to individual species and their attributes (Hill, 1979). The method is divisive since it involves the progressive refinement, by dichotomy, of a single axis through reciprocal averaging. In the TWINSpan algorithm, the dichotomy is originally located at the center of gravity (i.e., centroid) of the first ordination axis. The division is then refined using discriminant function analysis and weighting algorithms to place species/sites on one side of the dichotomy or the other. Each side of the dichotomy is then subdivided into smaller dichotomies using the same process. This division continues until some predefined criteria set by the user is reached (Jongman et al., 1995). Although the division is based on the species that contributes the most variation, the approach is polythetic because it uses information about several species simultaneously to perform this task. Finally, the method is hierarchical, producing a dichotomizing “tree” (dendrogram) of classes that can be considered at various levels of disaggregation. One of the common criticisms of TWINSpan is that dichotomies do not usually occur in nature, and sites are arbitrarily divided that could easily be grouped together. Final aggregations will often have inaccurate groupings of species/sites (Cumming, 2000; Hill, 1979; Jongman et al., 1995).

3.1.2. Correspondence analysis (CA)

Correspondence analysis maximizes dispersion of sites and species according to a latent environmental variable. One of the assumptions when performing CA on species data is that the species respond in a unimodal way to an unknown underlying latent variable(s). In this way, species occur within a limited range of each environmental variable. 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 the species scores along the first ordination axis and the correlation between species and sites. Multiple axes can be constructed, with the constraint that they are uncorrelated with the previous axes. With this method, it is not necessary to know the environmental cause of species distribution (i.e., the analysis can be performed on presence–absence data or abundance data). Correspondence analysis results can be represented on scatterplots, where a two-dimensional scatterplot will display the results of the first two axes, or the two most important latent environmental variables (Fig. 2). Here, the eigenvalue (λ) is a measure of how well the species scores correspond with the sample scores (Fig. 2). In particular, the eigenvalue of an axis will equal the correlation coefficient between species scores and sample scores.

Depending on the nature of the species data under study, CA can display two well-known mathematical faults, which are unrelated to any inherent structure in the data. First, site scores on the primary axis can be more compressed at the ends of the axis than in the middle,

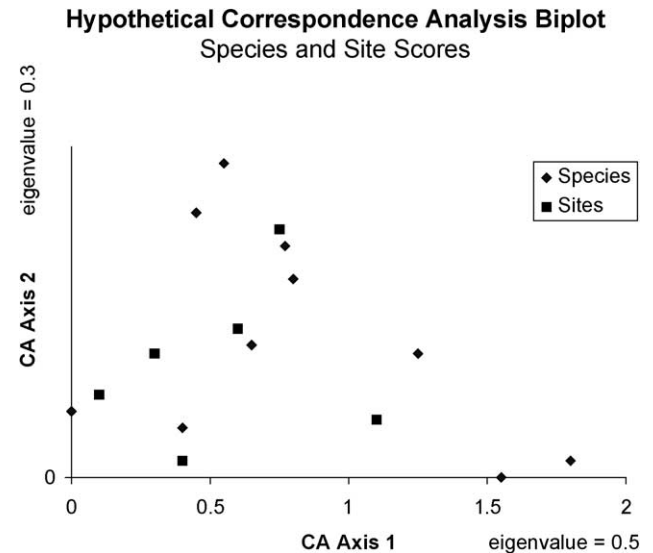


Fig. 2. Hypothetical correspondence analysis bi-plot. The eigenvalue (λ) is a measure of how well the species scores correspond with the sample scores. Specifically, the eigenvalue of an axis will equal the correlation coefficient between species scores and sample scores.

even if the change in species composition between sites is constant (Jongman et al., 1995). If care is not taken, this could cause an observer to attribute a stronger relationship between certain compressed sites than actually exists. Although the distance between site scores may be affected, this artifact does not impact the order of site arrangement along the axis. The second problem occurs when the second environmental gradient accounts for less variance than exists if the first axis is artificially folded onto itself. This will result in what is commonly termed the “arch effect”, which allows a second axis to be falsely created to maximize the dispersion along this folded axis (Jongman et al., 1995). This introduces false structure in the data by implying that the second axis contains new information about a latent variable. Hence, care must be taken when interpreting the structures depicted on the second axis. These structures may simply be artifacts that cannot be explained by real environmental conditions (Cumming, 2000).

3.1.3. Canonical correspondence analysis (CCA)

While CA maximizes dispersion of sites and species according to a latent environmental variable, canonical correspondence analysis (CCA) selects a linear combination of measured environmental variables to maximize the dispersion of sites and species (Jongman et al., 1995). In this manner, CCA is said to be a constrained form of CA, where there can be as many axes as there are environmental variables. CCA displays the main pattern in weighted averages for each species with respect to all selected environmental variables. The purpose is to detect patterns of variation in the data that can be best explained by the measured environmental variables, as well as to provide

an easily interpreted graphical representation of the species and environmental variables in low-dimensional space (Anderson & Clements, 2000; Cumming, 2000; Jongman et al., 1995).

3.1.4. Signature separability analysis

Another method of determining the natural arrangement of the plant communities within the peatland is through the examination of their spectral similarities and differences using spectral separability analysis (Anderson & Clements, 2000). This method is based purely on the spectral response of plant communities and makes the assumption that communities having a similar spectral response also have ecological similarities. The advantage of this technique is that the creation of spectrally separable groups should provide higher classification accuracies since most image classification algorithms apply hard decision rules to create discrete classes. However, the spectral response of vegetation communities is variable and complex, and groups that are ecologically significant and distinct may well have a similar spectral response. This fact accounts for many of the difficulties encountered by ecologists and remote sensing scientists when attempting to classify an ecosystem at a very high spatial resolution.

Swain and Davis (1978) describe three measures of separability: divergence, Jeffries–Matusita (J–M) distance, and transformed divergence. It has been shown that J–M distance analysis offers advantages when the data are normally distributed. The J–M distance algorithm calculates the separability of two class signatures, and outputs a value between 0 and 2 according to the following algorithm (Richards, 1993):

$$JM_{(i,j)} = 2[1 - e^{(-a_{(i,j)})}]$$

where $JM_{(i,j)}$ = Jeffries–Matusita distance between class i and j and $a_{(i,j)} = 0.125T[\mathbf{M}_{(i)} - \mathbf{M}_{(j)}]^* \text{Inv}[\mathbf{A}_{(i,j)}] * [\mathbf{M}_{(i)} - \mathbf{M}_{(j)}] + 0.5 \ln \{ \det(\mathbf{A}_{(i,j)}) / (\det(\mathbf{S}_{(i)}) * \det(\mathbf{S}_{(j)})) \}^{1/2}$, \mathbf{M} = mean vector; \mathbf{S} = covariance matrix.

Values approaching 2 indicate a high degree of separability, while those close to 0 indicate a low degree of separability. If sites have a low degree of separability, it is likely that there will be errors of omission/commission between them when using a parametric classifier (e.g., maximum likelihood). Hence, these distance measures can be used to determine how sites should be grouped according to their spectral separability. For instance, it is logical that sites demonstrating a high degree of separability (i.e., values approaching 2) represent different classes, while sites that have poor separability (i.e., values approaching 0) represent the same class, or at least classes that are spectrally similar. However, it should be noted that the relationship between separability and accuracy is nonlinear. Hence, care must be taken when interpreting separability based on these distance measures.

3.2. Data collection

The field component of the study involved quadrat sampling (each 3 m × 3 m) of 72 vegetation plots in 1995 (Fig. 3). Four of these sites fell outside of the CASI coverage and hence 68 sites were used for this analysis relating field measures to spectral reflectance. The majority of sample sites are located in the northern parts of the peatland due to an error in GPS sampling for sites in the southern portion. Vegetation plots were located within areas that were deemed ecologically similar and homogeneous. The sites sampled account for the ecological variability of the peatland and the spectral variability of the CASI data. This was evident in the field data and tested in the CASI data (i.e., the reflectance data for the 68 sites accounted for the spectral variability of the CASI image data, indicating that the spectral variations were sampled within the 68 sites).

The relevé method was employed, whereby the investigator is familiar with the vegetation and the existence of certain “community types” that appear to repeat themselves in similar habitats. This results in relatively discrete patches of similar vegetation. Within each plot, the species’ cover in nine subquadrants (1 m²) was recorded following the classification system of Daubenmire (1968). These values were then aggregated for a plot level cover estimate for each species. Using the Daubenmire method, the percent-cover categories were defined as follows: (1) 0–1%; (2) 1–5%; (3) 5–25%; (4) 25–50%; (5) 50–75%; (6) 75–95%; and (7) 95–100%. As in most relevé methods for estimating cover/abundance, the class intervals are unequal to allow for an easier estimation of a species-cover-to-area relationship than

Location of Vegetation Plots Within Fen Boundary

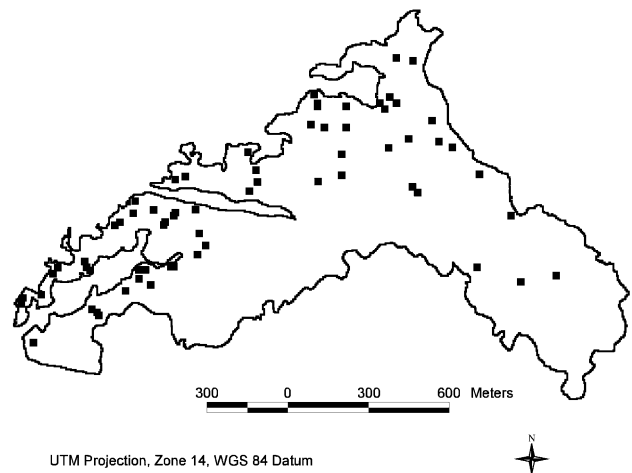


Fig. 3. Location of sample sites within the peatland complex. The field component of the study involved quadrat sampling (each 3 × 3 m) of 72 vegetation plots in 1995. Four of these sites fell outside of the CASI coverage and hence, 68 sites were used for the analysis relating field measures to spectral reflectance.

what is possible with equal intervals of cover. The rationale for having such a classification is that plant cover is very heterogeneous from point to point, thus making exact estimates rather problematic. The category intervals are most detailed at the low end, where differences in species cover matter most for classification. Also, the less abundant species with small cover may sometimes have an important diagnostic significance, hence requiring a more detailed breakdown of the lower scale values as compared to the larger scale values (Mueller-Dombois & Ellenberg, 1974). Field sampling catalogued over 114 different plant species in the peatland complex.

In 1995, GPS data were collected for the center of each vegetation plot using a Trimble Basic Plus receiver. Since a base station was not available, GPS data were post-processed using GPS PACE (GPS Positioning from ACS Clocks and Ephemerides), a program developed by the Canadian Geodetic Survey Division of Natural Resources Canada. GPS data were collected over a 30-min period at each site with positions recorded every 10 s. The overall root mean square error (RMSE) for the 68 vegetation plots was 2.52 and 2.08 m in the x (Easting) and y (Northing) directions, respectively.

To aid in the geo-referencing of the CASI data, 52 white reflectance targets (0.8×0.8 m) were placed across the peatland in July 1996 and their locations recorded with a Trimble Pro XL receiver. GPS data were again collected over a 30-min period, but with a measurement interval of 5 s. Data were again post-processed using GPS PACE. The RMSE for the 52 targets was 0.28 and 0.47 m in the x (Easting) and y (Northing) directions, respectively.

In an attempt to assess the absolute accuracy of the two different types of receivers, GPS data were collected at two survey monuments established by the Manitoba Department of Highways and Transportation, which had been surveyed in the spring of 1996. The control point coordinates were provided as UTM coordinates referenced to NAD83. The on-line version of the Canadian Geodetic Survey's "Geodetic Survey Routine: UTM/TM and Geographic" (GSRUG) was used to convert corrected positions from geographic to UTM coordinates. Based on measurements at the two survey monuments, the average total error of the Trimble Basic Plus and the Trimble Pro XL receivers was estimated to be 2.1 and 1.5 m, respectively. Average positions were determined using GPS PACE post-processing and then compared to the surveyed coordinates. These accuracy estimates and the manner in which vegetation plots were selected (i.e., within homogeneous patches) ensure appropriate geo-referencing for relating field-based measures to CASI reflectance data.

The CASI is a visible/near-infrared pushbroom imaging spectrograph with a reflection grating and a two-dimensional CCD solid-state array measuring 512×288 pixels (Gray et al., 1997). CASI data were acquired from a Piper Navajo Chieftan aircraft in July 1996 as part of the third BOREAS intensive field campaign-3 (IFC-3). Six spectral

channels were collected, sampling the visible and NIR portions of the electromagnetic spectrum (Table 1). These data were converted to radiance using software developed at the Center for Research in Earth and Space Technologies (CRESTech), using algorithms by Babey and Soffer (1992). The data were then converted to reflectance to eliminate atmospheric effects and compensate for changes in solar illumination during image acquisition (Gray et al., 1997). The flight lines were orthorectified and compiled into a mosaic to cover the entire peatland at a spatial resolution of 0.5×0.5 m.

3.3. Classification, error assessment, and cluster modification

Although there are a variety of algorithms developed for supervised classification, the maximum likelihood classification (MLC) technique was applied in this study. The MLC has become a standard classifier in remote sensing data analysis and has proven to be a robust algorithm in cases where the image data meet the assumptions required for parametric statistical analysis (Treitz, Howarth, Rotunno, & Soulis, 2000). To assess accuracy, the ground data were divided into calibration and validation sites for each class. The calibration sites were used to obtain statistics for the classification decision rules. The validation sites were used in the post classification error analysis, as an independent assessment of classification accuracy. Four measures of accuracy (i.e., overall accuracy, the confusion matrix, errors of commission/omission, and the Kappa coefficient (\hat{K})) were examined in the error analysis. Overall accuracy is a measure of the number of validation pixels that were classified correctly divided by the total number of validation pixels for all classes. 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 validation samples for that class in the validation set. The confusion matrix (also referred to as the error matrix or a contingency table) was used to illustrate class agreement and error in greater detail by illustrating the relationship between the independent validation sites (of a known class) and the percentage of these pixels actually classified into the various classes by the maximum likelihood classifier (Jensen, 1996; Lillesand & Kiefer, 2000). The percentages of pixels classified correctly are shown on

Table 1
CASI channel wavelengths

Channel	Waveband (nm) (FWHM)	Range
1	464–474	blue
2	525–536	green
3	549–600	green
4	631–650	red
5	743–753	NIR
6	853–864	NIR

FWHM = full-width half maximum.

the diagonal of the confusion matrix, while errors of commission (wrongful inclusion into class-row entries) and omission (wrongful exclusion from class-column entries) can be seen off the diagonal (Lillesand & Kiefer, 2000). Finally, Kappa coefficients 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 (Richards, 1993; Rosenfield & Fitzpatrick-Lins, 1986). Confidence intervals were calculated based on an assumed normal distribution of the \hat{K} statistic at a 95% level against a Type I error (Barber, 1989).

4. Results and discussion

4.1. Assessment of natural site arrangements

To determine how the species and sites were organized, CA was performed on the species abundance data. It was observed that species abundance varied between sites along the axis and the sites on opposite ends of the ordination axis

had no common species (i.e., species turnover). This suggested a unimodal response to latent environmental variables, which indicated that correspondence analysis (CA) was the most appropriate indirect ordination technique for these data (Cumming, 2000; Jongman et al., 1995; Thioulouse, Chessel, Doledec, & Olivier, 1997). In this case, CA was used to determine the natural arrangement of sites based on species presence and abundance, rather than to assess the impacts of latent variables. For this reason, it was only necessary to examine the primary ordination axis and group the sites based on their separations along this 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 clusters. The CCA used here incorporated 18 channels as environmental variables, which included the 6 original image channels as well as their standardized (correlation matrix) and nonstandardized (covariance matrix) principal components. This required the extraction of reflectance data for each site, in the form of the mean, standard deviation, and variance per band. A separate CCA analysis

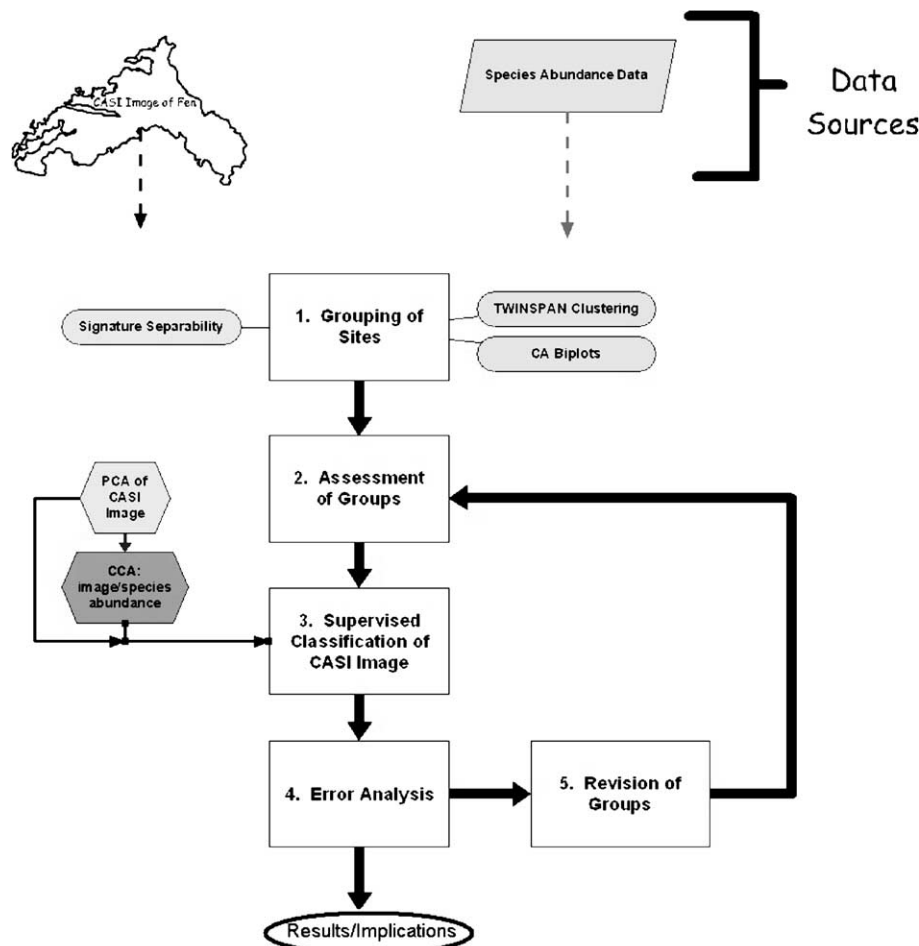


Fig. 4. Process chart outlining procedures used to create ecological classes using spectral criteria.

was performed for each of these measures. These values were then analyzed with the species abundance data using ADE-4 ordination software (Thioulouse et al., 1997).

Results of the CCA illustrated that the image channels and principal components did not relate well to the distribution of species abundance data with an explained variance of 44% and an r^2 value of 0.47. The eigenvalue for the first CCA axis was 0.21, less than half the first CA axis. The highest weightings for the canonical coefficients were for the first standardized and nonstandardized principal components. This indicates that these channels were the most significantly related to species distribution across sites. This approach varies from recent work completed by Andréfouët and Clereboudt (2000), who used similarity matrices to relate environmental variables to image spectra. Here, the goal was not to relate environmental variables to the image, but rather to determine how natural species clusters were described by spectral reflectance.

The CCA results could be considered the first opportunity to examine the raw ecological and reflectance data together. The results illustrated that the image channels and principal components did not relate to the distribution of species abundance data very well, suggesting that classification at the species level would be difficult. This finding supported the establishment of community groups, which may reduce some of the confusion introduced at the species level. Ideally, established community groups would incorporate species giving a similar or overlapping spectral response while maintaining a meaningful ecological relation. The canonical coefficients indicate that the first principal component was most closely related to species distribution across sites and supported the application of the first principal component as an input for the image classification.

4.2. Aggregation of sites using multivariate techniques

The procedures used to create the ecological classes, using spectral criteria, are described below and illustrated in Fig. 4.

(1) Initially, the sample plots were grouped into ecologically significant classes by using TWINSpan analyses of the species abundance data. The first division separated the sites into stands that were generally wet on the left side of the dendrogram and generally dry on the right. Subsequent divisions separated the peatland sites into eight classes (Table 2). The TWINSpan dendrogram and final groups used for classification are presented in Fig. 5.

(2) The CA ordination bi-plot was examined and sites were grouped based on separations along the primary axis (Fig. 6). This was done by creating group boundaries where there was a gap in the site scores of the first eigenvector, which can be shown numerically with the site scores, or graphically in horizontal space on the CA bi-plot. This created seven groups based on species abundance and composition (Table 3). Note that the seventh group is not

Table 2

TWINSpan classes with ecological descriptions

Code	Class	Description
1	<i>Warnstorfia exannulatus</i> / <i>Carex rostrata</i>	Extremely wet sites found in the northern and western regions of the peatland.
2	sedge fen	Nonwooded, open stands composed predominantly of sedges. In some areas, this class is a transition zone between <i>Warnstorfia exannulatus</i> lawns and shrub fens.
3	shrub fen	Characteristically dense shrub layer with no wooded canopy.
4	wooded fen	Open canopy of stunted <i>Larix laricina</i> present over a dense shrub layer.
5	<i>Scorpidium scorpioides</i> carpets	<i>Scorpidium scorpioides</i> and <i>Eriophorum alpinum</i> are indicator species.
6	forested fen	Occurs on drier sites with higher hummocks than wooded fens.
7	poor fen collapse	Associated with permafrost collapse scar.
8	<i>Sphagnum fuscum</i> bogs	Moss layer typically exceeds 75% cover.

shown in Fig. 6, but is composed of sites grouped at a larger ecological scale (poor fen collapse and *Sphagnum fuscum* bogs), which are evident in the larger-scaled bi-plot shown in Fig. 7.

(3) The J–M distance output was used to group the sites based on similarity of spectral response. Although somewhat arbitrary due to the nonlinear relationship between accuracy and separability, J–M distances were used as guidelines for determining the spectral separability of field sites (PCI, 1998; Schott, 1997). By using the separability measure for each site against all others, clusters were defined as follows:

- Sites found to have good separability (i.e., J–M distance > 1.9) from each other were not grouped together.
- If a site was shown to be poorly separable (i.e., J–M distance < 1.0) with another site, they 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 five initial classes based on spectral separability. The remaining 41 sites were shown in the separability analysis to be low to moderately separable (i.e., J–M distance > 1.0 and < 1.9). These sites were then added to the class from which they were most poorly separable. Group assignment was based on the following additional criteria:

- If a site had poor separability (i.e., J–M distance < 1.0) from any of the original clusters, it was added to that

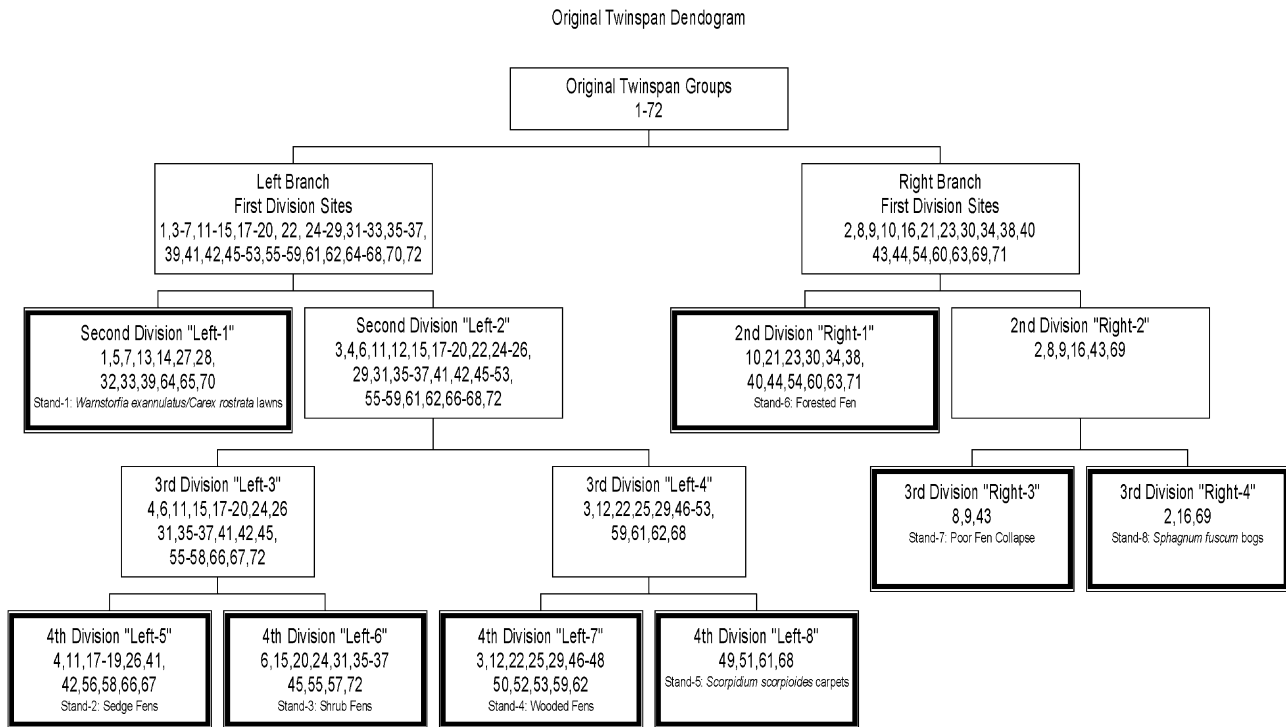


Fig. 5. Original TWINSpan dendrogram. The first division separates the sites into stands that were generally wet on the left side of the dendrogram and generally dry on the right. Subsequent divisions separated the peatland sites into eight classes (highlighted in bold). See Table 2 for class descriptions.

group, provided there were no conflicts with any member of that group. At this stage, there were still five classes, which were expanded one site at a time.

- If multiple sites had good separability (i.e., J–M distance >1.9) from all of the original clusters, a new class was created. This resulted in four new classes.

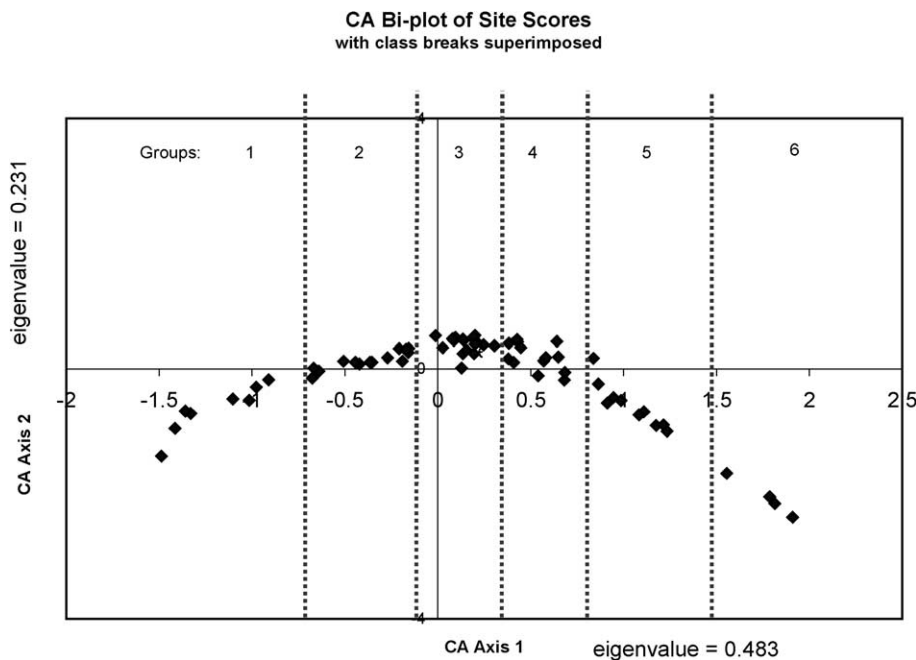


Fig. 6. Class divisions superimposed onto a correspondence analysis (CA) bi-plot. The CA ordination bi-plot was examined and sites were grouped based on separations along the primary axis. This was done by creating group boundaries where there was a gap in the site scores of the first eigenvector, which can be shown numerically with the site scores, or graphically in horizontal space on the CA bi-plot. This created six groups based on species abundance and composition. See Table 3 for class descriptions.

Table 3

Correspondence analysis classes with ecological descriptions and relations to TWINSpan classes

Code	Ecological descriptions
1	This group contains sites from TWINSpan group 6—forested areas. Forested peatlands are characterized by high hummocks of <i>Sphagnum warnstorffii</i> , <i>S. angustifolium</i> , and <i>Tomenthypnum nitens</i> , which support a canopy of <i>Larix laricina</i> with scattered <i>Picea mariana</i> .
2	This group is composed mainly of TWINSpan group 4, but also contains a site from each of TWINSpan groups 3 and 6. Indicator species include an open canopy of stunted <i>Larix laricina</i> , with <i>Tomenthypnum nitens</i> in the shrub layer, and <i>Aulacomnium palustre</i> on the hummocks.
3	Sites in this group are mainly from TWINSpan groups 3 and 4, but there is also one site from TWINSpan group 5. The area contains shrubs and stunted woody canopy.
4	This group is mainly composed of TWINSpan group 2, but also contains sites from TWINSpan group 1 and group 5. <i>Calliergonn giganteum</i> and <i>Menyanthes trifoliata</i> are the dominant species in this group.
5	This group has sites found in TWINSpan groups 1 and 2. These sites include <i>Warnstorffia exannulatus</i> and <i>Carex rostrata</i> lawns as well as nonwooded, open sedge areas.
6	Similar to TWINSpan group 1. <i>Warnstorffia exannulatus</i> and <i>Carex rostrata</i> are the most common species. These sites are extremely wet, associated with collapsed permafrost that is in contact with minerotrophic fen waters.
7	This group was comprised of the ecologically large-scaled sites, corresponding to groups 7 and 8 in the TWINSpan analysis, which are the poor fen collapses and <i>Sphagnum fuscum</i> bogs, respectively. Poor fen collapses are associated with permafrost collapse scars that are isolated from minerotrophic fen waters.

- Sites that had good separability (i.e., J–M distance >1.9) with all or most classes were not grouped (one site).
- Sites that could not be placed into any group without conflict were not grouped (four sites).

It can be seen that the above process resulted in nine J–M distance classes, with a 10th “unclassified” group containing five sites (Table 4).

Classes created based on CA ecological clusters and J–M distance spectral clusters were used to assess the suitability of TWINSpan image classification of the peatland complex. Each level of the TWINSpan dendrogram was examined, starting at the initial dichotomy. This was first done by examining the CA bi-plot to determine where on the first CA axis the first TWINSpan division occurred. One of the criticisms of TWINSpan is that sites close to but on opposite sides of each dichotomy will never be grouped together despite their ecological similarities (Cumming, 2000; Hill, 1979; Jongman et al., 1995; van Groenewoud, 1992). This leads to potential conflicts with spectral clusters as well. Therefore, J–M distance analysis was performed for sites on the primary CA axis that were also located near the first TWINSpan dichotomy. In terms of image classification, there would only be a significant conflict if sites were divided into different clusters when they were not spectrally separable. Conflicts that occurred at the first

dichotomy were carried on throughout all of the subsequent divisions. To determine the extent of the potential conflicts between spectral differences and TWINSpan clusters, each level of the hierarchy was examined.

4.3. Classification, error assessment, and cluster modification

Once initial clusters were created with the CA and J–M distance analysis, several maximum likelihood classifications were performed. As discussed, the main objective of this analysis was to incorporate the TWINSpan and/or CA clusters in the image classification, in hopes of achieving a higher classification accuracy than that which was achieved by any of these means alone, while maintaining ecologically meaningful classes. Once the classification accuracies for the individual methods were determined, the results were examined in detail to create more suitable classes. This involved the examination of the commission and omission errors in the confusion matrices to determine the extent that validation pixels were incorrectly classified.

Ecological classes derived from the ordination analysis with high classification accuracy were left unchanged. Within these clusters, individual sites that were identified through J–M distance analysis to be potential conflicts were removed from the cluster and analyzed separately. The remaining sites were those that were grouped together ecologically, but had significant spectral conflicts. These included sites that were close in proximity on the CA axis, but were separated in the TWINSpan divisions. Here, the CA bi-plots were very helpful to regroup these sites without enforcing hard decision rules. Sites that were not spectrally separable from an ecological cluster and that were located close to the cluster on the CA axis were merged. Separability analyses were performed each time class sites were adjusted in order to determine the spectral impact of these minor adjustments on the ecological clusters. This logic was applied in an iterative fashion to group all sites (including the ecologically large-scaled sites) in an ecologically meaningful way that reduced spectral conflicts as much as possible. The optimal set of classes, along with a description of contributing techniques for the derivation of each class, is provided in Table 5.

The locations of the TWINSpan classes (‘poor fen collapse’ and ‘*S. fuscum* bogs’) are superimposed on the CA bi-plot in Fig. 7. These two classes are very distinctly separated from the remainder of the sites on the CA bi-plot. This is a result of a very unique species composition for these sites. Although TWINSpan does succeed in separating these sites from the remainder of the peatland, the dendrogram in Fig. 5 does not portray any indication of the uniqueness of these classes. However, the CA bi-plot illustrates that there are two scales of natural ecological clustering occurring in the peatland complex. Three groups can be seen at a coarse scale (‘poor fen collapse’,

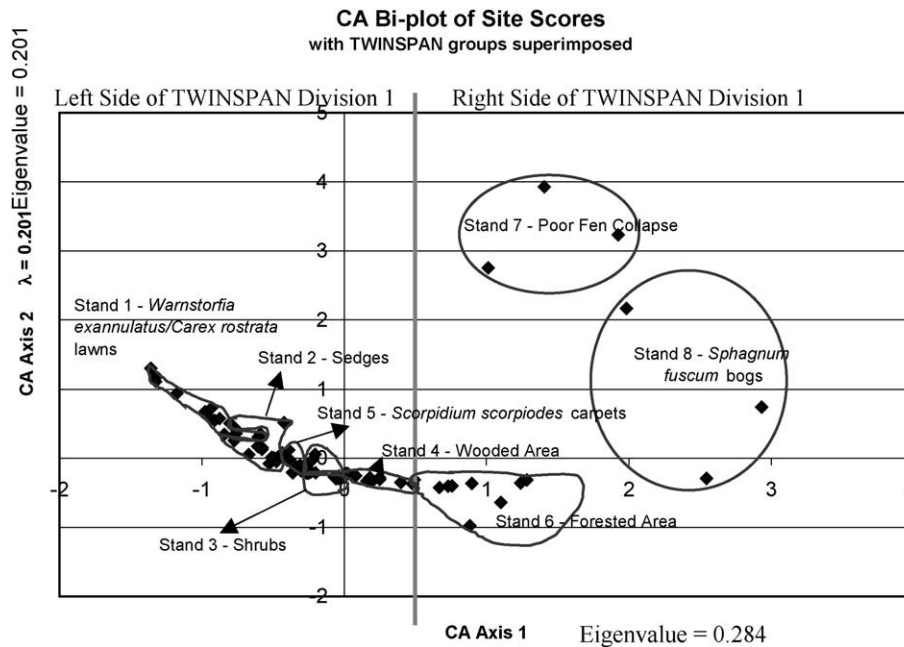


Fig. 7. Original correspondence analysis (CA) bi-plot with TWINSpan groups superimposed. However, the CA bi-plot illustrates that there are two scales of natural ecological clustering occurring in the peatland complex. Three groups can be seen at a coarse scale ('poor fen collapse', '*S. fuscum* bogs', and 'all other sites'), each containing a distinct collection of species/abundance. However, to obtain a more detailed classification of the peatland, it is necessary to further subdivide the 'all other sites' cluster into ecologically meaningful groups.

'*S. fuscum* bogs', and 'all other sites'), each containing a distinct collection of species/abundance. However, to obtain a more detailed classification of the peatland, it is necessary to further subdivide the 'all other sites' cluster into ecologically meaningful groups. Fig. 7 reveals that it is within these subgroups that ecological conflicts can be seen between the TWINSpan dichotomies and the CA bi-plot.

Conflicts resulting from the first TWINSpan dichotomy can be seen by comparing Figs. 5 and 8. The CA results identified a problem with the TWINSpan placement of site 71, which grouped more closely with the left side of the TWINSpan division than the right. Note that sites 71 and 62 are located in very close proximity to one another (in terms of eigenvector space), but that site 62 fell under the left side of the TWINSpan division, while 71 was grouped with the right side. Despite their obvious similarities, these sites will never be grouped together with the TWINSpan clustering. Examination of the superimposed TWINSpan groups in Fig. 7 also illustrates similar conflicts with the 2nd, 3rd, and 4th TWINSpan dichotomies, between Stands 1–6 and the natural site arrangement shown on the CA bi-plot.

4.4. Image classification—TWINSpan classes

Maximum likelihood classification showed very poor agreement for classes derived from TWINSpan clustering, with an overall accuracy of 41% and a Kappa coefficient of 0.32. Examination of the confusion matrix (Table 6)

revealed considerable variation in class accuracy (ranging from 0% to 78%), with considerable commission and omission errors. The previously discussed ecological conflicts evident at the first TWINSpan division were also apparent in the confusion matrix. There were problems with the TWINSpan clusters even at the first division, which were illustrated by errors of commission and omission between classes 1–5 (left side of TWINSpan division) and 6–8 (right side of TWINSpan division). This was particularly notable for the forest (Class 6) (right side of TWINSpan division), wooded (Class 4) (left side of TWINSpan division), and shrub areas (Class 3) (left side of TWINSpan division) (Table 2). For example, the confusion matrix showed that 32.4% of the forest validation pixels were classified as wooded area and 17.6% as shrubs, suggesting similarity in spectral response for these sites. This agreed with the results found in the CA bi-plot, which illustrated that forested, wooded, and shrub areas (identified by TWINSpan) were very similar in terms of species abundance and distribution. The implication of this example is that TWINSpan is not a suitable method of clustering species abundance data to optimize image classification for this environment.

4.5. Image classification—correspondence analysis (CA) classes

Maximum likelihood results for classes derived from correspondence analysis were also very low, with an overall accuracy of 42% and a Kappa coefficient of 0.40 (Table 7).

Table 4
Jeffries–Matusita spectral classes with ecological descriptions

Code	Ecological descriptions
1	This group has some similarities with group 4. The most dominant species include <i>Hamatocaulis vernicosus</i> and <i>Calliergon giganteum</i> , with <i>Menyanthes trifoliata</i> , <i>Carex limosa</i> , and <i>Carex diandra</i> also common.
2	<i>Scorpidium scorpioides</i> and <i>Eriophorum alpinum</i> are indicator species for this vegetation association where both species reach their highest cover values. <i>Menyanthes trifoliata</i> is also present, though not in as high abundance as in group 3.
3	<i>Sphagnum fuscum</i> is the indicator species of this group, dominating the moss layer with cover values typically greater than 75%. <i>Sphagnum angustifolium</i> , which occurs as scattered patches in wet depressions, occurs with regularity. A scattered shrub layer of <i>Kalmia polifolia</i> and <i>Chamaedaphne calyculata</i> is characteristic, with <i>Larix groenlandicum</i> being present on sites supporting <i>Picea mariana</i> . Other common species with low abundance include <i>Vaccinium oxycoccos</i> , <i>Carex aquatilis</i> , and <i>Rubus chamaemorus</i> .
4	These sites are extremely wet, associated with collapsed permafrost that is in contact with minerotrophic fen waters. Indicator species for this group are <i>Carex rostrata</i> and <i>Warnstorfia exannulatus</i> , which both occur with high frequency and abundance. <i>Carex limosa</i> , <i>C. chordorrhiza</i> , <i>Utricularia intermedia</i> , and <i>Menyanthes trifoliata</i> are also major components of this vegetation type. The mosses <i>Hamatocaulis vernicosus</i> and <i>Calliergon giganteum</i> , and the vascular plants, <i>Salix pedicularis</i> , <i>Comarum palustre</i> and <i>Carex diandra</i> are frequently present in regions transitional to drier sedge fens. <i>Scorpidium scorpioides</i> occurs locally with high abundance.
5	This group is composed mainly of nonwooded open areas with sedges predominant in the herbaceous layer. On wetter stands, submerged to emergent <i>Calliergon giganteum</i> predominates in the moss layer, while drier stands support firm carpets of <i>Calliergon giganteum</i> and <i>Hamatocaulis vernicosus</i> . <i>Menyanthes trifoliata</i> is most dominant in this group. <i>Comarum palustre</i> and the sedges <i>Carex limosa</i> , <i>Carex chordorrhiza</i> , <i>Carex lasiocarpa</i> , and <i>Carex diandra</i> are common to abundant.
6	This is a shrub group. Indicator species include <i>Betula nana</i> , <i>Campyllum stellatum</i> , <i>Bryum pseudotriquetrum</i> , <i>Brachythecium mildeanum</i> , <i>Aulacomnium palustre</i> , <i>Hypnum lindbergii</i> , and <i>Tomenthypnum nitens</i> . A dense shrub layer of <i>Betula nana</i> , <i>Salix pedicularis</i> , and <i>Andromeda polifolia</i> , with no wooded canopy, is characteristic of this vegetation type.
7	<i>Betula pumida</i> , <i>Carex limosa</i> , and <i>Hamatocaulis vernicosus</i> are common species in this group.
8	This group includes an open canopy of stunted <i>Larix laricina</i> over a dense shrub layer of <i>Betula nana</i> and <i>Chamaedaphne calyculata</i> . <i>Sphagnum warnstorfii</i> and <i>Vaccinium oxycoccos</i> occur with high frequency on flat hummocks. <i>Equisetum fluviatile</i> is one of the common vascular plants occurring in the hollows.
9	This is the largest of the J–M groups, with 16 sites. The indicator species that separate this group from others include <i>Larix laricina</i> , <i>Tomenthypnum nitens</i> , and <i>Aulacomnium palustre</i> . Microtopographic relief is greater than in other areas, with <i>Tomenthypnum nitens</i> , <i>Sphagnum warnstorfii</i> , and <i>Aulacomnium palustre</i> forming flat hummocks interspersed with wet hollows. <i>Hypnum pratense</i> is common on the sides of hummocks and as low mounds in hollows.

Errors of commission/omission were high for all classes, with individual class accuracies less than 45% for all classes. Results show that vegetation classes derived solely from species and abundance data are not suitable for spectral classifiers. This is due to the similarity in spectral response for many species, and the within-species variations that result in a lack of a unique spectral response. For example, seasonal variations, health/stress, age, and nutrient regime will affect the spectral response (and potentially even localized evolution) within a given species (Okin & Roberts, 2000). Also, as with any ordination technique, it is difficult to define discriminant class boundaries along a gradient as derived from this type of analysis of ecological data.

4.6. Image classification—J–M classes

Results obtained with classes derived solely from signature separability analysis showed a marked improvement from results obtained with TWINSpan and CA classes, with an overall accuracy of 60% and a Kappa coefficient of 0.52 (Table 8). Errors of commission and omission were greatly reduced with this technique, and were highest in off-diagonal classes closest to the diagonal. In other words, confusion occurred along class divisions where sites were moderately separable (i.e., similar spectrally and ecologically). Classes that were very spectrally distinct from each other were not confused in the maximum likelihood classifier, resulting in commission/omission errors of 0% for these cases. For example, Classes 5 and 3 (Table 4) were spectrally very different and no confusion occurred between these two classes. Examination of the species abundance data for the sites that made up these classes revealed that the classes were very different in terms of dominant species (Table 4). For example, the species *Menyanthes trifoliata* was highly abundant in Class 5 (between 50% and 95% coverage for most sites), but not present at all in Class 3. In contrast, there is significant confusion in class assignment between Classes 3 and 4. The species abundance data indicate that sphagnum species in Class 3 are also present at most of the Class 4 sites, although not in at the same abundance. Both Classes 3 and 4 are described as being very wet, with characteristics associated with permafrost scars. The similarity in species and in the moisture regime causes considerable spectral overlap between the sites, making it difficult to separate classes based solely on spectral response. A by-product of grouping classes in this manner is that they tend to be “ecologically complex” as evidenced by their ecological descriptions (Table 4).

4.7. Image classification—“optimized” classes

The “optimized” classification represents an attempt to incorporate the species abundance data that was organized by CA analysis into the signature separability analysis. The best final overall classification accuracy for the optimally derived clusters was 62% ($\hat{K}=0.55$) (Tables 9 and 10),

Table 5
Grouping criteria and ‘Optimized’ classes with ecological descriptions

Optimal classes	Sites	Dominant technique(s)	Comments	Ecological description
1	21, 38, 40, 60, 63, 2	TWINSpan	TWINSpan Stand 6 and ecologically large-scaled site 2	dominantly forested
2	10, 23, 34, 44, 54, 71, 3, 12, 22, 25, 48, 52, 62, 29, 47, 57	CA	Parts of TWINSpan Left 2 and Right 1. Could not be derived from TWINSpan, but sites are located closely together on the first CA axis.	combination of forested and wooded communities
3	17, 19, 66, 67, 36, 31	TWINSpan J–M distance	From TWINSpan Left 3.	combination of sedges and shrubs
4	1, 13, 14, 55, 65	TWINSpan J–M distance	most from TWINSpan Left 1 (55 from Left 6)	dominantly <i>Warnstorfia exannulatus</i> / <i>Carex rostrata</i> lawns
5	43, 61, 69, 16, 9	J–M distance	Not derived with TWINSpan or CA. Sites from both sides of the first TWINSpan division.	combination of poor fen collapse, <i>Sphagnum fuscum</i> bogs, and <i>Scorpidium scorpioides</i> carpets
6	18, 42, 50, 51, 53, 68, 49	TWINSpan J–M Distance	From TWINSpan Left 2	combination of sedges, wooded communities, and <i>Scorpidium scorpioides</i> carpets
7	26, 39, 56, 64	TWINSpan J–M distance CA	From TWINSpan Left 1 and Left 5.	combination of <i>Warnstorfia exannulatus</i> / <i>Carex rostrata</i> lawns and sedges
8	4, 11, 28, 32, 46, 72	TWINSpan J–M distance CA	From TWINSpan Left 1 and Left 2	combination of <i>Warnstorfia exannulatus</i> / <i>Carex rostrata</i> lawns, sedges, shrubs, and wooded communities

which represents a 21% improvement over the TWINSpan classification, and a slight but not statistically significant improvement over the signature separability classes. Note that the relatively low \bar{K} coefficient 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. Again, there is a wide range of class accuracies (ranging from 33% to 94%), with the highest accuracy occurring for a group composed of a combination of sedges and shrubs. These sites were located in clusters separated by the TWINSpan dichotomies, again showing that TWINSpan is not a suitable tool to derive classes for supervised image classification. However, the CA bi-plot was found to

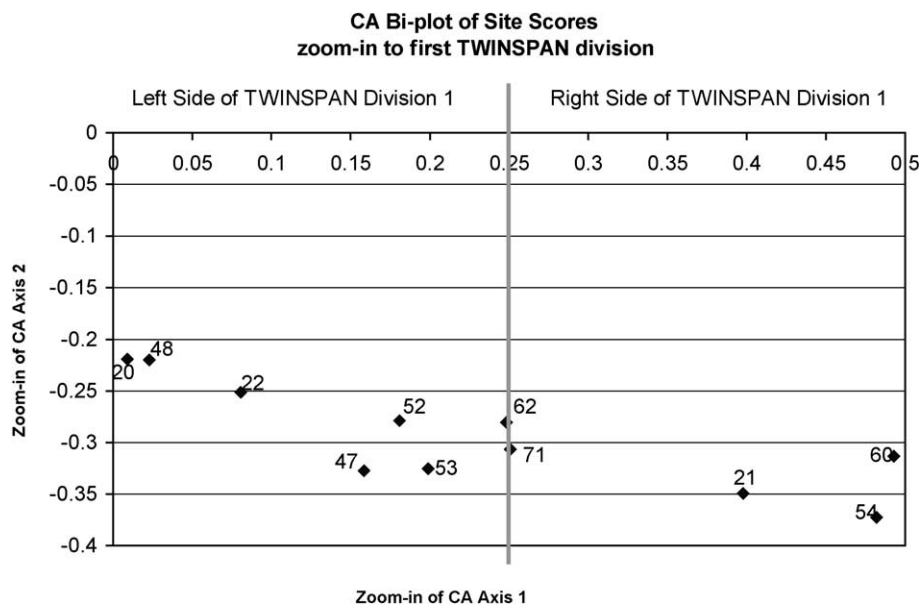


Fig. 8. Illustration of potential conflicts arising in TWINSpan analysis as depicted on the CA bi-plot. Conflicts result from the first TWINSpan dichotomy, which separates ecologically similar sites 62 and 71. Note that sites 71 and 62 are located in very close proximity to one another, but that site 62 fell under the left side of the TWINSpan division, while 71 was grouped with the right side. Despite their obvious similarities, these sites will never be grouped together with the TWINSpan clustering. Examination of the superimposed TWINSpan groups in Fig. 7 also illustrates similar conflicts with the 2nd, 3rd, and 4th TWINSpan dichotomies between Stands 1–6 and the natural site arrangement shown on the CA bi-plot.

Table 6

Confusion matrix for TWINSpan classification

Reference class	Number of pixels	Percent classified into class							
		1	2	3	4	5	6	7	8
1	180	40.0	22.6	11.1	7.9	36.1	1.9		2.8
2	252	16.7	26.2	9.4	4.6			5.6	
3	180	10.0	39.3	48.9	18.5	2.8	17.6		
4	216	13.3	10.3	25.6	51.4	30.6	32.4		
5	72	13.3	1.2		2.3	30.6			
6	216			5.0	14.8		47.2		97.2
7	36	1.1						77.8	
8	36	5.6	0.4		0.5		0.9	16.7	0.0

Overall accuracy: 41.2%; Kappa coefficient: 0.317; 95% confidence interval = 0.291–0.343.

be helpful in deriving this class because it highlighted similarities in species composition that allowed signature separability groups to be appropriately modified and gave the final class ecological meaning.

It was 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 clusters 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 versus validation sites for the classification accuracy assessment. One would expect a source of error if the various ecological clusters were not represented evenly in both the calibration and validation data sets.

Ecologically large-scaled sites (i.e., those with very different species composition) were very problematic for grouping purposes. Both TWINSpan and CA results indicated that they could be classified together, but 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 ecologically large-scaled sites into other clusters.

4.8. Suitability of the techniques and possible alternatives for improved accuracy

Overall, results showed poor classification accuracy using groups defined with purely ecological techniques

Table 7

Confusion matrix for correspondence analysis (CA) classification

Reference class	Number of pixels	Percent classified into class					
		1	2	3	4	5	6
Unclassified		16.7	16.7				
1	216	39.8	10.7	5.0			
2	216	39.4	30.6	13.5	0.5	1.1	
3	288	1.9	12.5	34.4	21.8	19.4	
4	216	1.9	14.4	36.1	37.0	26.1	
5	180		7.4	9.4	40.7	43.3	
6	108	0.5	7.9	1.7		10.0	100.0

Overall accuracy = 42.2%; Kappa = 0.402; 95% confidence interval = 0.365–0.439.

Table 8

Confusion matrix for Jeffries–Matusita (J–M) classification

Reference class	Number of pixels	Percent classified into class								
		1	2	3	4	5	6	7	8	9
1	72	30.6					11.1	6.5		0.9
2	108		39.8	27.8	33.3		0.7		2.8	8.3
3	72	12.5	19.4	47.2	7.4					0.9
4	108		2.8	25.0	59.3					1.2
5	108	22.2				63.9	23.6			3.1
6	144	27.8				36.1	42.4	9.3		0.9
7	108	6.9	21.3				9.0	81.5		9.0
8	144		2.8						89.6	14.5
9	324		13.9				13.2	2.8	7.6	61.1

Overall accuracy = 59.6%; Kappa = 0.519; 95% confidence interval = 0.496–0.542.

(TWINSpan and CA). Improved accuracies for some classes were achieved when consideration was given to the spectral signatures of the sites. However, confusion was evident for all techniques when attempting to characterize classes with indeterminate boundaries (i.e., related closely to environmental/species gradients which are not distinct). This “blending” of class boundaries is evident on the CA bi-plots (Figs. 7 and 8), and suggests that, at the scale of this study, species composition and abundance data will not sufficiently characterize a peatland environment to allow it to be spectrally classified by species associations. Fuzzy classification techniques may prove helpful to characterize these indeterminate boundaries. Here, fuzzy calibration class weights would be assigned based on known species mixtures within a site. During the classification, each pixel is assigned a membership grade with respect to the fraction of the pixel belonging to each class (Lillesand & Kiefer, 2000). The output would then contain some homogeneous pixels (for example, a membership grade of 1.0 for “shrubs” and 0.0 for all other classes) as well as mixed pixels in the areas where boundaries are not distinct (for example, a pixel could be assigned a membership grade of 0.4 for shrubs, 0.3 for forest, 0.2 for wooded areas, 0.1 for sphagnum, and 0.0 for all other classes).

Larger-scale plant communities are well suited for a fuzzy classification scheme, but the small-scale processes

Table 9

Confusion matrix for “optimized” classification

Reference class	Number of pixels	Percent classified into class							
		1	2	3	4	5	6	7	8
1	108	61.1	17.0				3.7		
2	288	33.3	64.6		1.4		27.8	1.4	2.8
3	108		6.9	93.5			4.6	36.1	4.6
4	72	0.9			55.6	29.2	7.4		
5	72		0.7		19.4	54.2			
6	108	4.6	4.9	1.9	16.7	16.7	33.3		10.2
7	72			4.6	2.8		0.9	51.4	9.3
8	108		5.9		4.2		22.2	11.1	73.2

Overall accuracy = 62.4%; Kappa coefficient = 0.553; 95% confidence interval = 0.527–0.579.

Table 10
Comparison of classification results

Original grouping method	Classification specifications	Overall accuracy (%)	\hat{K} lower (95%)	Kappa coefficient (\hat{K})	\hat{K} upper (95%)
TWINSpan	maximum likelihood, 6 image channels, no null class	41.2	0.291	0.317	0.343
CA	maximum likelihood, 6 image channels + standardized PC-1, no null class	42.2	0.365	0.402	0.439
J–M distance	maximum likelihood, 6 image channels, no null class	59.6	0.496	0.519	0.542
Optimal clusters	maximum likelihood, 6 image channels, no null class	62.4	0.527	0.553	0.579

and species groups evident in a fen community may reduce the practicality of applying such a scheme to a fen environment. To date, fuzzy techniques have had most success when applied to coarser resolution imagery (i.e., AVHRR, Landsat TM, and SPOT), where subpixel components are clearly identifiable (e.g., Atkinson, Cutler, & Lewis, 1997; Cross, Settle, Drake, & Paivinen, 1991; Huguenin, Karaska, Van-Blaricom, & Jensen, 1997). Further, although fuzzy techniques would potentially provide a detailed classification, an accuracy assessment of the output would be very challenging on a per-pixel level. It might prove helpful to consider the underlying environmental gradients that are known to be present in bi-plot environments, such as moisture, pH, and water chemistry. These gradients could be incorporated as a layer in the classification and also used to better group the species abundance data (i.e., incorporated into ecological clustering/ordination techniques such as canonical correspondence analysis). Further, environmental gradients could be used to assign a class weight that varies spatially (for example, as moisture increases, the probability that the pixels would belong to a wet class, such as sites dominated by *Warnstorfia exannulatus* and *Carex rostrata*, would also increase).

5. Conclusions

Based on the results discussed above, a number of generalizations can be made regarding the relationship between species abundance, distribution, and spectral characteristics for high spatial resolution CASI imagery. Examination of the techniques commonly used by ecologists to define species associations based on presence and abundance illustrated that these methods alone were not optimal for classification of the peatland at very high spatial resolutions. First, known problems with TWINSpan clustering were evident at all levels of the TWINSpan divisions, which were compounded when classes were subjected to spectral classification. At all stages of the analysis, there were vegetation sites that were separated by a TWINSpan division that were not spectrally separable. As a result, classification results from the original TWINSpan clusters had the lowest classification accuracy of all methods tested. This confirms that this technique for ecological grouping is

not readily amenable to remote sensing data analysis. It was observed, however, that the most spectrally separable TWINSpan classes occurred at multiple levels within the TWINSpan divisions/dendrogram. These results seem logical, in that it would be expected that spectral groupings would not correspond on a one-to-one basis with TWINSpan classes at a single level within the hierarchy.

Correspondence analysis by itself also proved to be ineffective in defining groups suitable for image classification. However, CA did prove to be useful when examined in conjunction with remote sensing techniques such as signature separability analysis. It was demonstrated that an integrated application of CA bi-plots, TWINSpan clustering, and J–M analysis provided slightly 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 classification of these types of environments when using high spatial resolution data.

The difficulties encountered in delineating classes along indeterminate boundaries suggest that increased classification accuracies will be difficult to achieve without incorporation of driving environmental variables. These environmental variables and their gradients across space should provide insight into the function and processes of the peatland communities, which would help to define ecologically meaningful associations that can be spectrally delineated. It is recommended that future work with high spatial resolution data in peatland communities incorporate these data to further combine ecological and remote sensing techniques.

Acknowledgements

The authors thank Jim Freemantle for preprocessing the CASI data at the Centre for Research in Earth and Space Technology (CRESTech) and Bob Steinauer for assistance in vegetation sampling and species identification. This work was supported by grants from NASA BOREAS program to Dennis Jelinski, the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of a Special Collaborative Research Grant to J. Harry McCaughy and Peter Lafleur, and NSERC Research Grants to Paul Treitz and Dennis Jelinski.

References

- Anderson, M. J., & Clements, A. (2000). Resolving environmental disputes: a statistical method for choosing among competing cluster models. *Ecological Applications*, 10(5), 1341–1355.
- Andréfouët, S., Claerebout, M. (2000). Objective class definitions using correlation of similarities between remotely sensed and environmental data. *International Journal of Remote Sensing*, 21(9), 1925–1930.
- Atkinson, P. M., Cutler, M. E. J., Lewis, H. (1997). Mapping sub-pixel proportional land cover with AVHRR imagery. *International Journal of Remote Sensing*, 18(4), 917–935.
- Babey, S., & Soffer, R. (1992). Radiometric calibration of the Compact Airborne Spectrographic Imager (CASI). *Canadian Journal of Remote Sensing*, 18(4), 233–242.
- Barber, D. G. (1989). KAPPA: a Macintosh program to compare levels of agreement between image classifications and ground reference data. Waterloo, Ontario, Canada: Institute for Space and Terrestrial Science, University of Waterloo (ISTS-EOL-TR89-006).
- Bubier, J. L., Crill, P. M., Moore, T. R., Savage, K., Varner, R. S. (1998). Seasonal patterns and controls on net ecosystem CO₂ exchange in a boreal peatland complex. *Global Biogeochemical Cycles*, 12, 703–714.
- Bubier, J. L., Moore, T. R., Juggins, S. (1995). Predicting methane emission from bryophyte distribution in northern Canadian peatlands. *Ecology*, 76, 677–693.
- Bubier, J. L., Moore, T. R., Roulet, N. T. (1993). Methane emissions from the mid-boreal region in northern Ontario, Canada. *Ecology*, 74, 2240–2254.
- Bubier, J. L., Rock, B. N., Crill, P. M. (1997). Spectral reflectance measurements of boreal wetland and forest mosses. *Journal of Geophysical Research*, 102(29), 483–494.
- Calmé, S., Desrochers, A. (2000). Biogeographic aspects of the distribution of bird species breeding in Québec's peatlands. *Journal of Biogeography*, 27, 725–732.
- Cross, A. M., Settle, J. J., Drake, N. A., Paivinen, R. T. M. (1991). Sub-pixel measurement of tropical forest cover using AVHRR data. *International Journal of Remote Sensing*, 12(5), 1119–1129.
- Cumming, B. (2000). Department of Biology, Queen's University. Personal Communication.
- Daubenmire, R. F. (1968). Plant communities: a textbook of plant synecology. New York: Harper and Rowe.
- Elliot-Fisk, D. (2000). The taiga and boreal forest. In M. G. Barbour, & W. D. Billings (Eds.), *North American terrestrial vegetation* (2nd ed.) (pp. 41–73). Cambridge, UK: University of Cambridge Press.
- Gauch, H. G. Jr. (1982). *Multivariate analysis in community ecology*. Cambridge, UK: Cambridge Univ. Press.
- Gauch, H. G. Jr., Whittaker, R. H. (1981). Hierarchical classification of community data. *Journal of Ecology*, 69, 537–557.
- Glaser, P. H. (1992). Peat landforms. In H. E. Wright Jr., B. Coffin, & N. Aaseng (Eds.), *The patterned peatlands of Minnesota* (pp. 3–14). Minneapolis: University of Minnesota Press.
- Gorham, E. (1991). Northern peatlands: role in the carbon cycle and probable responses to global warming. *Ecological Applications*, 1, 182–195.
- Gray, L., Freemantle, J., Shepherd, P., Miller, J., Harron, J., Hersom, C. (1997). Characterization and calibration of the CASI airborne imaging spectrometer for BOREAS. *Canadian Journal of Remote Sensing*, 23(2), 188–195.
- 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*. New York: Microcomputer Power.
- Huguenin, R. L., Karaska, M. A., Van-Blaricom, D., Jensen, J. R. (1997). Subpixel classification of bald cypress and tupelo gum trees in thematic mapper imagery. *Photogrammetric Engineering and Remote Sensing*, 63(6), 717–725.
- Jacobsen, A., Nielson, A., Ejrnaes, R., Groom, G. B. (1999). Spectral identification of Danish grassland classes related to management and plant species composition. *Proceedings of the Fourth International Airborne Remote Sensing Conference and Exhibition/21st Canadian Symposium on Remote Sensing, June 21–24, vol. 1* (pp. 74–81). Ann Arbor, USA: ERIM International Inc., Ottawa, Ontario.
- Jensen, J. R. (1996). *Introductory digital image processing: a remote sensing perspective*. Upper Saddle River, NJ: Prentice-Hall.
- Jongman, R. H. G., Ter Braak, C. J. F., van Tongeren, O. F. R. (1995). *Data analysis in community and landscape ecology*. Cambridge, UK: Cambridge Univ. Press.
- Lafleur, P. M., McCaughey, J. H., Joiner, D. W., Bartlett, P. A., & Jelinski, D. E. (1997). Seasonal trends in energy, water, carbon dioxide fluxes at a northern boreal wetland. *Journal of Geophysical Research*, 102(D24), 29009–29020.
- Lavoie, C., Saint-Louis, A. (1999). The spread of gray birch (*Betula populifolia* Marsh.) in eastern Québec: landscape and historical considerations. *Canadian Journal of Botany*, 77, 859–868.
- Lavoie, C., Zimmermann, C., & Pellerin, S. (2001). Peatland restoration in southern Québec (Canada): a paleoecological perspective. *Écoscience*, 8, 247–258.
- Lillesand, T., & Kiefer, R. (2000). *Remote sensing and image interpretation*. (4th ed.). New York: Wiley.
- Ludwig, J. A., & Reynolds, J. F. (1988). *Statistical ecology: a primer on methods and computing*. New York: Wiley.
- Mausel, P. W., Kamber, W. J., & Lee, J. K. (1990). Optimum band selection for supervised classification of multispectral data. *Photogrammetric Engineering and Remote Sensing*, 56(1), 55–60.
- Mazerolle, M. J., Drolet, B., & Desrochers, A. (2001). Small-mammal responses to peat mining of southeastern Canadian bogs. *Canadian Journal of Zoology*, 79, 296–302.
- Mitch, W. J., & Gosselink, G. J. (1993). *Wetlands*. (2nd ed.). New York: Van Nostrand-Reinhold.
- Moore, T. R., Heyes, A., & Roulet, N. T. (1994). Methane emissions from wetlands, southern Hudson Bay lowland. *Journal of Geophysical Research*, 99, 1455–1467.
- Mueller-Dombois, D., & Ellenberg, H. (1974). *Aims and methods of vegetation ecology*. New York: Wiley.
- National Wetlands Working Group (NWWG) (1988). *Wetlands of Canada*. Ecological Classification Series, 24, Environment Canada, Montreal (452 pp.).
- Okin, G. S., & Roberts, D. A. (2000). Linear unmixing of simulated, noisy spectra: vegetation detection limits in areas of low cover. In *Proceedings of the 2000 AVIRIS Airborne Geoscience Workshop*. Pasadena, CA: Jet Propulsion Laboratory (10 pp.).
- PCI (1998). *Imageworks version 6.3 EASI/PACE*. Richmond Hill, Canada: PCI Geomatics.
- Petrone, R. M., Waddington, J. M., & Price, J. S. (2001). Ecosystem scale evapotranspiration and net CO₂ exchange from a restored peatland. *Hydrological Processes*, 15, 2839–2845.
- Potter, C., Bubier, J., Crill, P., & Lafleur, P. (2001). Ecosystem modeling of methane and carbon dioxide fluxes for boreal forest sites. *Canadian Journal of Forest Research*, 31, 208–223.
- Richards, J. A. (1993). *Remote sensing digital image analysis*. New York: Springer-Verlag.
- Rosenfield, G. H., & Fitzpatrick-Lins, K. (1986). A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 52(2), 223–227.
- Roulet, N. T. (2000). Peatlands, carbon storage greenhouse gases, and the Kyoto Protocol: prospects and significance for Canada. *Wetlands*, 20, 605–615.
- Schlesinger, W. H. (1997). *Biogeochemistry: an analysis of global change*. London: Academic Press.
- Schott, J. R. (1997). *Remote sensing: the image chain approach*. Toronto: Oxford Univ. Press.
- Sellers, P. J., Hall, F. G., Ransom, J., Margdis, H., Kelly, B., Baldocchi, D., den Hartog, J., Cihlar, J., Ryan, M., Goodison, B., Crill, P., Ranson, K. J., Lettenmaier, D., & Wickland, D. E. (1995). The Boreal Ecosystem–Atmosphere Study (BOREAS): an overview and early results from the

- 1994 field year. *Bulletin of the American Meteorological Society*, 76, 1549–1577.
- Swain, P. H., & Davis, S. M. (1978). *Remote sensing: the quantitative approach*. New York: McGraw-Hill.
- Thioulouse, J., Chessel, D., Doledec, S., & Olivier, J. M. (1997). ADE-4: a multivariate analysis and graphical display software. *Statistics and Computing*, 7, 75–83.
- Treitz, P. M., & Howarth, P. J. (2000). Integrating spectral, spatial, and terrain variables for forest ecosystem classification. *Photogrammetric Engineering and Remote Sensing*, 66(3), 305–317.
- Treitz, P. M., Howarth, P. J., Rotunno, O., & Soulis, E. D. (2000). Agricultural crop classification using SAR tone and texture statistics. *Canadian Journal of Remote Sensing*, 26(1), 18–29.
- Treitz, P. M., Howarth, P. J., & Suffling, R. C. (1992). Application of detailed ground information to vegetation mapping with high spatial resolution digital imagery. *Remote Sensing of Environment*, 42, 65–82.
- van Groenewoud, H. (1992). The robustness of correspondence, detrended correspondence, and TWINSpan analysis. *Journal of Vegetation Science*, 3, 239–246.
- Waddington, J. M., & Roulet, N. T. (1996). Atmosphere–wetland carbon exchanges: scale dependency of CO₂ and CH₄ exchange on the developmental topography of a peatland. *Global Biogeochemical Cycles*, 10(2), 233–245.
- Waddington, J. M., & Warner, K. D. (2001). Atmospheric CO₂ sequestration in restored mined peatlands. *Ecoscience*, 8, 359–368.
- Whiting, G. J. (1994). CO₂ exchange in the Hudson Bay lowlands: community characteristics and multispectral reflectance properties. *Journal of Geophysical Research*, 99, 1519–1528.
- Wright, H. E. Jr., Coffin B. A., & Aesang N. E. (Eds.). (1992). *The patterned Peatlands of Minnesota*. Minneapolis: University of Minnesota Press.