INT. J. REMOTE SENSING, 2002, VOL. 23, NO. 21, 4761–4776

Taylor & Francis
Taylor & Francis Group

A multivariate approach to vegetation mapping of Manitoba's Hudson Bay Lowlands

R. K. BROOK

Natural Resources Institute, University of Manitoba, Winnipeg, Manitoba, R3T 2N2, Canada; e-mail: umbrook1@cc.umanitoba.ca

and N. C. KENKEL

Department of Botany, University of Manitoba, Winnipeg, Manitoba, R3T 2N2, Canada; e-mail: kenkel@cc.umanitoba.ca

Abstract. The Hudson Bay Lowlands of Manitoba contain a wide range of vegetation types that reflect local variations in climate, geological history, permafrost, fire, wildlife grazing and human use. This study, in Wapusk National Park and the Cape Churchill Wildlife Management Area, uses a Landsat-5 TM image mosaic to examine landscape-level vegetation classes. Field data from 600 sites were first classified into 14 vegetation classes and three unvegetated classes. Principal component analysis was used to examine the spectral properties of these classes and identify outliers. Multiple discriminant analysis was then applied to determine the statistical significance of the vegetation classes in spectral space. Finally, redundancy analysis was used to determine the amount of vegetation variance explained by the spectral reflectance data. We advocate this adaptive learning approach to vegetation mapping, by which the researcher employs an iterative strategy to carefully examine the relationship between ground and spectral data. This approach is labour intensive, but has the advantage of producing vegetation classes that are spectrally separable, decreasing the likelihood of errors in classification caused by overlap between classes.

1. Introduction

Remotely sensed data from satellites have been used for thematic land cover mapping at a wide range of spatial scales and for numerous applications (Cihlar 2000). Such data have most commonly been used to interpret land cover that reflects the combined characteristics of numerous landscape components. Continental and global scale projects have typically produced generalized landscape level maps that identify broadly defined land cover classes (e.g. 'agricultural', 'grassland', 'boreal forest' or 'barren land') that are separable without having to resort to detailed ground cover information. However, such broadly defined map classes are often of limited utility in addressing landscape-scale ecological questions. In response to this limitation, finer-scale maps have been produced in which land cover is related to identifiable ground cover components (Roughgarden *et al.* 1991). For example, vegetation maps

This paper was presented at the 6th Circumpolar Symposium on Remote Sensing of Polar Environments held in Yellowknife, Northwest Territories, Canada, from 12–14 June 2000.

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160110113917

represent plant communities based on their floristic composition and structure. Many vegetation mapping projects have used a standard land cover mapping approach, applying a single supervised or unsupervised classification directly to the entire image and subsequently performing an accuracy assessment. In many cases, such maps have low accuracy and therefore could not be used for their intended purpose (Townshend 1992).

Reasons for the unfulfilled potential of satellite remote sensing in vegetation mapping are numerous. A primary limitation of satellite sensor data is that they provide only a measure of the amount and type of reflected and emitted electromagnetic energy. This information is affected by a number of factors, including soil moisture, substrate, plant structural configuration, topography and atmospheric effects, as well as the amount, vigour, productivity, structure and floristic composition of vegetation (Richardson and Wiegand 1977, Vogelmann and Moss 1993). The complexity of ecosystem structure results in a continuum of vegetation cover and a corresponding continuum in multi-spectral space (Richards 1993). Subdivision of this multi-spectral continuum into meaningful vegetation classes is a major challenge that requires careful consideration.

Spectral classification is essentially a modelling procedure (*sensu* Jeffers 1982) whereby a complex continuum of land cover information is translated into ecologically meaningful classes. Vegetation mapping is an important practical exercise, since a level of generalization is required if ecosystem complexity is to be represented in a meaningful and readily interpretable way. In the initial verification of a classification model, reflected spectral data are related to ground cover information. Specifically, field-collected vegetation data are examined to establish the feasibility of using satellite imagery to separate known ground cover classes (Cihlar 2000). The developed classification model is then normally applied to the entire area of interest in the satellite image to produce a vegetation map. This is followed by model validation, in which map accuracy is assessed using an independently derived ground-survey data set. All phases of model development, including verification and validation, are highly dependent on the quality and quantity of ground-survey data. Detailed site-level information on vegetation composition and structure is required in order to objectively define meaningful ground cover classes and provide essential insights into factors determining spectral reflectance.

Spectral classification is necessarily a subjective process that involves carefully considered sequential decisions. The quality and utility of a spectral classification is determined by the analyst's skill, judgement and familiarity with the study area (Foody 1999), as well as the scale of consideration. Each mapping project has unique challenges since the concept of spatial heterogeneity is a scale-dependent descriptor of the inter-relatedness of land cover components across the landscape (O'Neill *et al.* 1988). Decisions must be made regarding the sampling design used to collect ground cover data, including the number and distribution of sample sites. The ground cover data must then be analysed and the sites allocated to mutually exclusive vegetation classes based on floristic composition (Matveyeva 1994, Legendre and Legendre 1998). Following this, the defined vegetation classes must be translated into spectrally separable classes. This is a critical step, since there is no guarantee that floristically distinct land cover classes will have distinct spectral signatures. Indeed, an accurate vegetation map assumes that each land cover class has unique and characteristic spectral properties over at least part of the measured electromagnetic spectrum. Classes lacking unique spectral signatures must either be amalgamated (e.g. Matthews 1991) or separated using ancillary data (e.g. Nilsen *et al.* 1999).

A number of studies have examined the relationship between spectral reflectance and broad-scale land cover classes (e.g. Perry and Lautenschlager 1984, Hope *et al.* 1993, Korobov and Railyan 1993, Cihlar *et al.* 1997, Homer *et al.* 1997, Nemani and Running 1997). However, this relationship has not been as well studied for finer-scale vegetation classes, particularly those based on floristic composition and structure. Determining the relationship between spectral reflectance and vegetation composition is particularly important in northern ecoregions, where vegetation mapping presents unique logistic and theoretical challenges. Much of the arctic and sub-arctic is inaccessible, making the collection of ground cover data difficult and expensive. Arctic ecosystems are spatially and floristically heterogeneous and vegetation may be sparse or non-existent in many areas (Ferguson 1991, Nilsen *et al.* 1999). Arctic vegetation is characterized by low shrubs, graminoids, bryophytes and lichens (Bliss *et al.* 1973). Lichens, which have unique spectral reflectance properties (Petzold and Goward 1988), often form the dominant ground cover. A number of approaches have been used to classify and map northern vegetation in the Hudson Bay Lowlands (Ritchie 1962, Pala and Boissonneau 1982) and other northern regions (Thompson *et al.* 1980, Horn 1981, Matthews 1983, Shasby and Carneggie 1986, Ferguson 1991, Matthews 1991, Pearce 1991, Morrison 1997, Nilsen *et al.* 1999), with varying degrees of success.

Bivariate scatterplots (comparing two-band combinations of spectral data) are commonly used to examine class separability and to determine the ideal band combinations for classification purposes (e.g. Ferguson 1991, Muller *et al.* 1999). While this approach is useful for comparing band pairs, vegetation classification generally involves the simultaneous use of three or more spectral bands. Multivariate analysis is a more optimal strategy for simultaneously comparing the spectral reflectances of numerous satellite bands (Richards 1993). In multivariate analysis, interrelationships among the original variables (e.g. multiple spectral bands) are summarized as a reduced set of derived variables (Legendre and Legendre 1998). Multivariate analysis thus provides a powerful strategy for data verification, such as comparing the separability of vegetation classes and examining the utility of different band combinations in characterizing classes.

In this paper, we use ground information and spectral data from multiple Landsat-5 Thematic Mapper (TM) bands to develop an iterative strategy for the classification of highly complex sub-arctic vegetation in northern Manitoba, Canada. Our objective is to demonstrate the utility of multivariate methods in developing and verifying a vegetation mapping approach based on ground cover data and remotely sensed spectral information. A complete overview of the entire classification process is beyond the scope of this paper. Instead, emphasis is placed on the examination of vegetation class separability in order to maximize accuracy of the final map product. The multivariate techniques considered include cluster analysis, principal component analysis (PCA), correspondence analysis (CA), multiple discriminant analysis (MDA) and redundancy analysis (RDA).

2. Study area

The study area includes Wapusk National Park and much of the Cape Churchill Wildlife Management Area, which collectively encompass a large portion of the Hudson Bay Lowlands of Manitoba, Canada (figure 1). The area is a flat, extensive coastal plain that forms a broad transition zone between continuous boreal forest to the south and arctic tundra to the north. Plant communities of the Hudson Bay

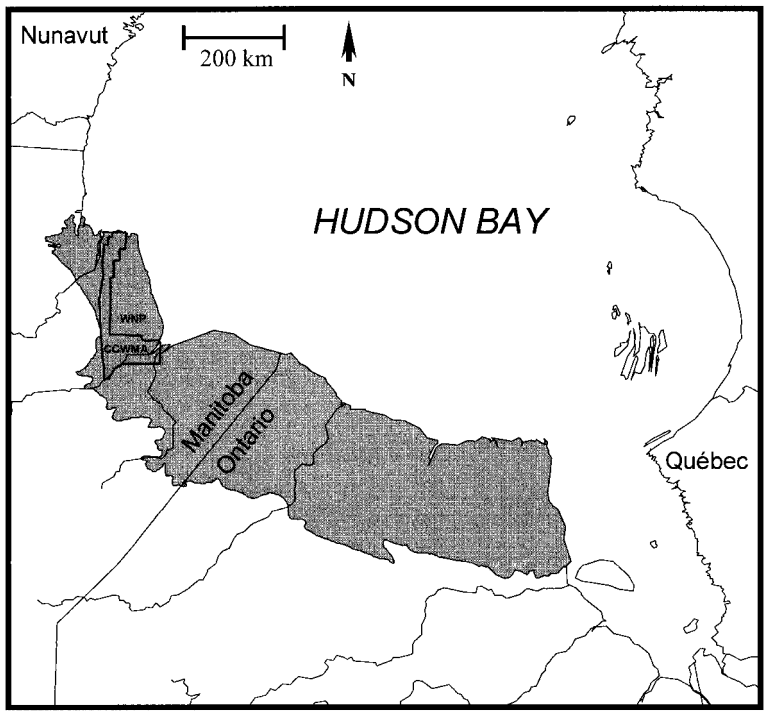

Figure 1. Wapusk National Park (WNP) and the Cape Churchill Wildlife Management Area (CCWMA) situated within the Hudson Bay Lowlands, Canada (grey shading).

Lowlands include coastal salt marshes, upland heaths and extensive fen and bog complexes of highly variable tree cover intermixed with vast numbers of ponds and lakes (Ritchie 1956, Sims *et al.* 1982, Pala and Weischet 1982). Vegetational variation follows a spatio-temporal gradient inland from the coast, reflecting ongoing isostatic emergence of the area from Hudson Bay (Ritchie 1962, Webber *et al.* 1970). Grazing by large herbivores, including geese and caribou, has significantly altered plant community structure in some areas (Campbell 1995, Ganter *et al.* 1996). Fire is a frequent and recurring process on the landscape. The dynamic nature of the region, together with its size and remote location, makes satellite imagery an essential tool for vegetation mapping. The complex mosaic of vegetation in the region provides an ideal case study for examining issues of class separability in landscape-level vegetation mapping.

3. Methods

3.1. *Analytical approach*

The collection and analysis of ground cover and spectral reflectance data, and classification of the satellite image, followed an iterative approach that included a number of decision steps (figure 2). Our methodology focused on making careful decisions based on multivariate data analysis and the inherent limitations of satellite imagery. Adaptive learning was an essential component of our procedure, with each decision having an important impact on subsequent decisions in the mapping process. Cluster analysis (Ward's method, Legendre and Legendre 1998), based on a chord distance matrix of ground cover data, was performed to delineate major vegetation

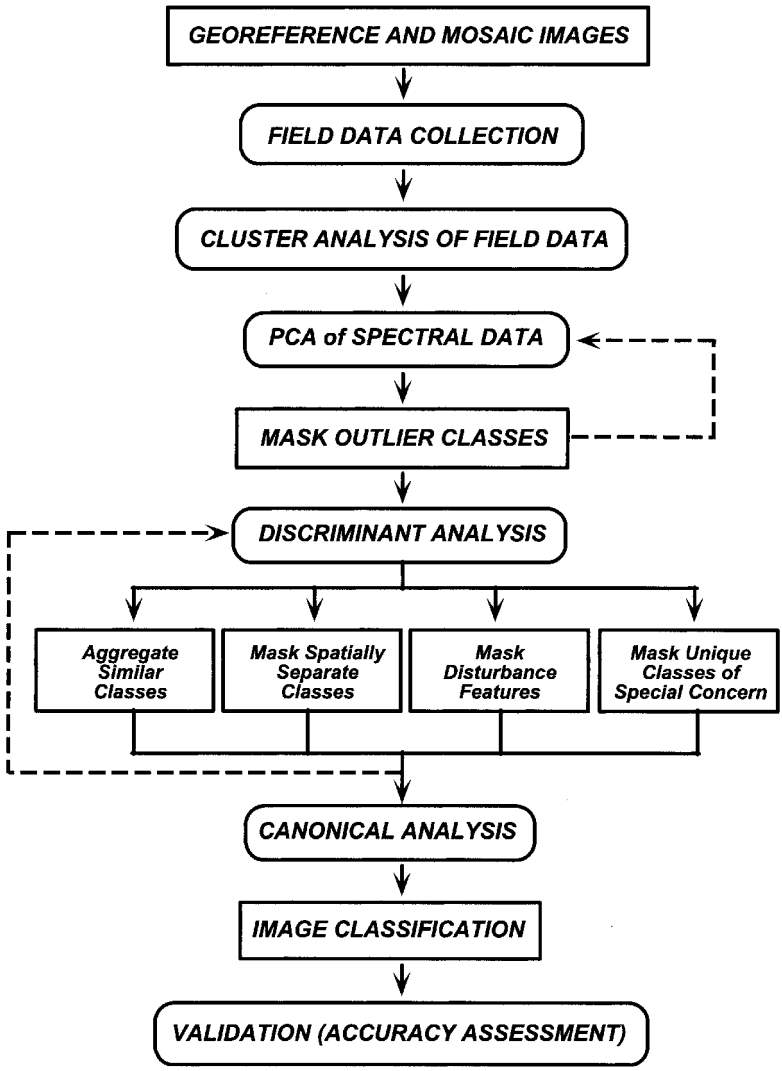

Figure 2. Flow model diagram of the generalized approach to satellite image classification using multivariate data analysis. Square boxes are steps involving satellite image manipulation and classification, while rounded boxes are data analysis steps. Dashed arrows depict iterative steps.

communities (HIERCLUS, Podani 1994). The spectral data of the defined vegetation classes were then analysed using principal component analysis, based on a correlation matrix (ORDIN, Podani 1994). This step identifies classes that have highly unique spectral signatures in order to classify them separately and better resolve differences between classes that are more similar. The separability of individual vegetation classes was then assessed using multiple discriminant analysis (MDA) (ORDIN, Podani 1994). MDA maximizes the ratio of the between-to within-groups variance, and graphically displays interclass variation in a low-dimensional ordination space (Legendre and Legendre 1998). MDA thus provides valuable insight into the relative separability of vegetation classes. Finally, the correlation between the floristic and

spectral data was examined in order to verify the model relating ground vegetation cover to spectral data in the satellite imagery. Canonical analysis is the appropriate model to examine the correlation between two multivariate data sets (Gittins 1985). We used a canonical model known as redundancy analysis (RDA) to examine the relationship between spectral reflectance (TM bands 3, 4, 5 and 7) and vegetation ground cover. RDA is an extension of multiple regression for modelling multivariate response data (Legendre and Legendre 1998). The method constrains the vegetation data such that ordination vectors are linear combinations of spectral reflectance values.

3.2. *Data collection*

Two full Landsat TM images covering the study area, taken on 27 July 1996, were acquired. These were the most recent cloud-free TM scenes available for the study area during the peak of the growing season (early July to mid-August). All non-thermal bands were included. The thermal band (TM6) was not used because of its low spatial resolution (120 m). In order to remove geometric distortion in the Landsat TM imagery, each scene was geocoded to a Universal Transverse Mercator (UTM) grid referenced to the North American Datum of 1927 (NAD 27) zone 15 projection. A total of 15 ground control points (GCPs) were collected for each image (GCPWORKS, PCI Geomatics 1998) from 1:50 000 NAD 27 National Topographic System (NTS) paper maps that are the most accurate spatial data currently available for the region. GCP selection was based on the ability to find a corresponding pixel within the image and the maps accurately. Three GCPs were selected that were common to both images in order to ensure that when mosaicked together, the fit would be accurate. Image to image GCP collection, first order transformation and nearest neighbour re-sampling of uncorrected imagery were performed. Image identifiable points were selected and matched to map coordinates. The green, red and near-infrared bands of the Landsat TM image displayed as a false composite image were used for point identification because it produced maximum contrast between land, vegetation and water features. The resulting Root Mean Square (RMS) error for the imagery was 0.96 pixel (28.8 m), 0.63 pixel (18.9 m) (x, y) for the north image and 0.57 pixel (17.1 m), 0.87 pixel (26.1 m) (x, y) for the south image. Once the two scenes were georeferenced, they were mosaicked together to produce a single composite image. This geocoded image was then subset to size it to the bounds of the study area. A mask was digitized over the ice in Hudson Bay and it was cut out of the image to remove its confusing effect on the classification process.

Field data were collected throughout the study area from June to September of 1998 and 1999. Four hundred sites that were visually homogenous at a minimum scale of 50 m × 50 m were sampled. At the centre of each site, percent cover values for each plant species and unvegetated substrate were estimated within a 10 m × 10 m plot. Each plot was then sub-sampled using two randomly placed 2 m × 2 m plots. The precise location of each site was determined using a hand held Global Positioning System (GPS) by taking several location fixes per second for 30–45 minutes and then averaging them over time. An additional 200 sites were independently sampled for classification accuracy assessment. The location data were used to obtain Landsat 5 spectral reflectance values (TM bands 1–5 and 7) for all sites, which allowed us to directly compare the ground cover information to the thematic data provided by the satellite image.

3.3. *Initial class allocation*

Cluster analysis provided an objective grouping of vegetation communities, independent of the spectral data, resulting in fourteen vegetation types (table 1). Three unvegetated classes (< 20% vegetation cover) were also identified based on prominent physical characteristics: water, unvegetated ridges and unvegetated shorelines. Digital numbers (DN) of the six Landsat TM bands were obtained for each sample site to form a verification data set. Pixels were removed from the dataset if they clearly fell outside of the normal range of homogenous sites. This step was performed to ensure that the training data set contained no mixed pixels (Arai 1992). For illustration purposes, a representative sample of 15 pixels was randomly selected from each of the 17 classes in order to indicate inter-class spectral variability.

3.4. *Identification of outlier classes*

The spectral data of the defined vegetation classes (17 classes, each with 15 sites and six Landsat bands) were analysed using principal component analysis, based on a correlation matrix. PCA is a multivariate ordination method that considers intercorrelations of the spectral data to produce an optimized and simplified representation of the underlying data structure. PCA is particularly well suited to the identification of outlier groups in multidimensional spectral band space. In our analyses, sites are represented as scores on the two major component axes, while the six spectral bands are shown as ordination biplot scores (Gabriel 1971). Classes appearing as prominent outliers in the two-dimensional ordination space were identified and removed. Classes were considered 'outliers' when the cluster of 15 sites defining the class was clearly separated from the remaining sites. After one or more outlier classes were removed, PCA was run again on the reduced data set. This process was repeated until no strong outlier groups remained.

Table 1. Land cover and vegetation classes of the study area with the average digital number from Landsat TM imagery (s.d.); $n = 15$ for each class.

	Landsat TM Digital Number (DN)					
Class	TM1	TM2	TM3	TM4	TM5	TM7
1. Sphagnum Larch Fen	55 (1.4)	24 (0.7)	22 (1.1)	75 (2.2)	49 (1.5)	16 (1.4)
2. Sedge Rich Fen	55 (1.4)	23 (1.0)	20 (1.1)	81 (3.6)	59 (2.2)	18 (1.3)
3. Willow Birch Shrub	53 (0.7)	21 (0.5)	18 (0.6)	64 (2.4)	58 (3.0)	19 (1.8)
4. Sedge Larch Fen	55 (2.1)	21 (0.8)	20 (1.3)	46 (2.0)	50 (2.8)	18 (1.2)
5. Sphagnum Spruce Bog	52 (1.3)	22 (0.5)	18 (0.5)	47 (2.1)	39 (1.7)	13 (0.7)
6. Graminoid Willow Salt Marsh	65 (4.6)	31 (3.9)	34 (7.2)	64 (7.8)	70 (6.3)	25 (2.6)
7. Willow Sedge Poor Fen	57 (2.1)	25 (1.6)	24 (1.6)	62 (4.8)	60 (9.5)	21 (3.2)
8. Lichen Spruce Bog	63 (2.7)	28 (1.3)	29 (2.2)	63 (4.6)	68 (3.5)	25 (2.0)
9. Sedge Bullrush Poor Fen	59 (1.9)	24 (1.1)	26 (1.1)	51 (2.1)	69 (3.2)	26 (1.6)
10. Lichen Melt Pond	61 (2.5)	25 (0.9)	27 (1.8)	45 (3.6)	72 (4.4)	29 (2.5)
11. Lichen Peat Plateau	73 (1.3)	33 (0.9)	39 (1.4)	59 (1.6)	90 (1.3)	35 (1.1)
12. Dryas Heath	70 (10.2)	31 (6.9)	35 (9.3)	55 (4.9)	99 (10.3)	42 (8.3)
13. Regenerating Burn	56 (1.8)	23 (0.9)	23 (1.8)	42 (3.5)	70 (1.7)	30 (2.4)
14. Recent Burn	53 (0.7)	20 (0.5)	21 (1.0)	24 (1.7)	38 (1.9)	24 (1.9)
15. Unvegetated Ridge	122 (12.6)	68 (8.5)	89 (11.8)	86 (11.0)	150 (19.5)	83 (9.9)
16. Unvegetated Shoreline	87 (6.1)	45 (4.8)	56 (8.4)	53 (6.7)	54 (4.7)	23 (2.6)
17. Water	48 (2.8)	17 (4.1)	14 (3.1)	7 (0.7)	4 (1.6)	3 (1.1)

Based on results of the iterative PCA, three major divisions of the data were identified: vegetated, unvegetated and water. An unsupervised 203-class isodata classification was run on the entire image mosaic (IMAGEWORKS, PCI Geomatics 1998), and the resulting classes assigned to one of these three divisions. The number of classes selected for the isodata unsupervised classification was based on the number of vegetation classes being separated and the size of the area being classified. From the unsupervised classification, the unvegetated group was separated into either the unvegetated ridge class (high reflectance in TM bands) or unvegetated shoreline class (low reflectance in TM bands). The water, unvegetated ridge and unvegetated shoreline classes were then converted to individual bitmap masks in order to remove them from further image classification.

The boundaries of known recent burns were digitized to create a bitmap mask over the recently disturbed areas. An unsupervised 28-class isodata classification was then run on the image, and all classes within the spectral range of recent burns were aggregated to form a single recent burn class. This was converted into a bitmap mask and removed from further classification.

3.5. *Discriminating vegetation classes*

Once the main outlier groups were removed from the spectral data set, MDA was used, based on spectral bands 3, 4, 5 and 7 identified as the optimal band set in PCA analysis. In our analysis, the vegetation classes are represented as 95% confidence ellipsoids, while the spectral bands are shown as biplot scores (Gabriel 1971). MDA was used to determine whether the floristically-based vegetation classes overlapped in spectral band space. Overlapping classes were then carefully examined to determine the nature of their spectral similarity. For each class, a decision was made to do one of the following: (1) Retain the class, while acknowledging that overlap with another class will compromise the accuracy of the final classification; (2) Aggregate two overlapping classes, if it can be established that they are su^{ffi}ciently similar floristically and if doing so does not compromise the utility of the final map; (3) Mask a class, if it is a disturbance feature that crosses over vegetation communities and is spatially distinct; (4) Mask a class, if it is floristically distinct and can be spatially separated from its overlapping class; (5) Mask a class, if it contains features that are considered unique and important to the final map and might be lost during classification and filtering.

3.6. *Verifying the model*

Once overlapping classes and strong outliers are removed, more subtle relationships between spectral reflectance and vegetation cover on the ground can be examined. Floristic data were obtained from 10 sites in each of the nine vegetation classes that were identified as spectrally separable in the MDA analysis. A total of 55 plant species and three unvegetated variables (water, organic soil and plant litter) were included. These data were first processed using correspondence analysis (CA) ordination (ORDIN, Podani 1994). This step was necessary to linearize the data and to reduce its dimensionality (Green 1993). RDA was then performed on the 90 sites, using the scores on the first four CA ordination axes as the response variables and the four spectral reflectance bands as the explanatory variables.

4. Results

4.1. *Principal component analysis*

PCA ordination of all 17 classes (table 1) indicated two strong outliers, water and unvegetated ridge classes (figure 3(*a*)). The water class is characterized by low reflectance in all bands, while spectral band reflectances for the unvegetated ridge class are uniformly high. A residual PCA (water and unvegetated ridge classes removed) revealed an additional outlier class, Unvegetated Shoreline (figure 3(*b*)). This class is characterized by very high reflectance in the visible spectrum (bands 1–3). Following removal of the Unvegetated Shoreline class, the Recent Burn class proved to be an outlier with low reflectance in all bands (figure 3(*c*)). PCA of the remaining 13 classes suggested no additional strong outliers (figure 3(*d*)).

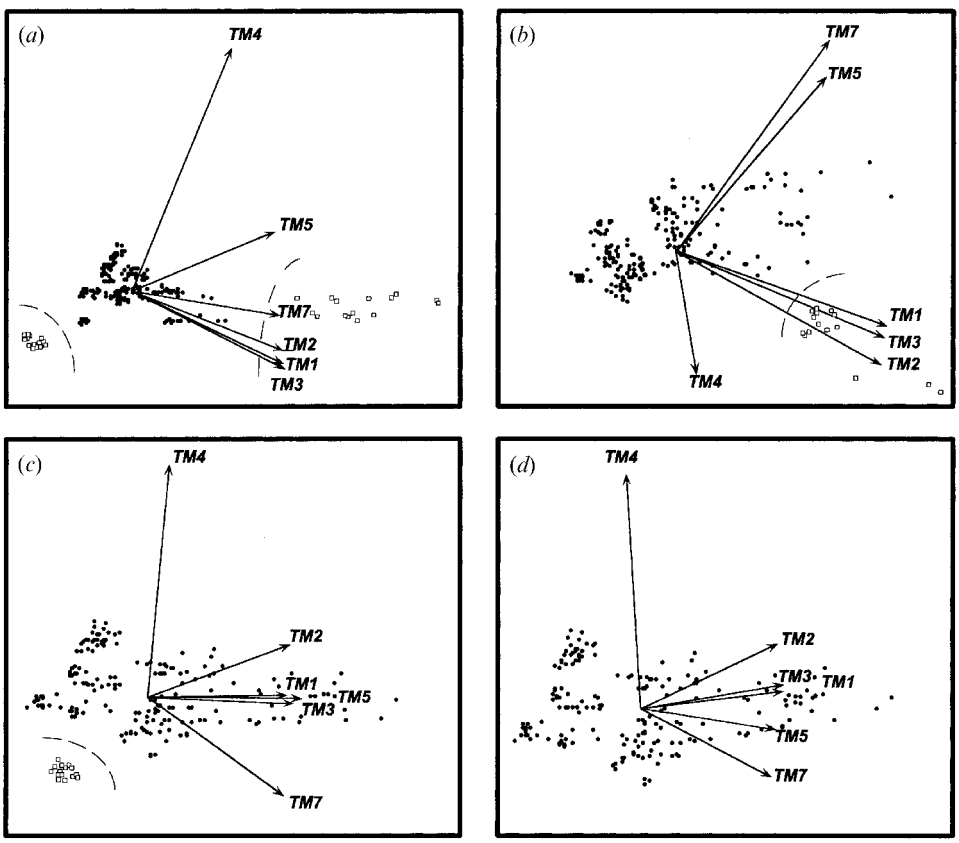

Figure 3. Principal component analysis of spectral reflectance variables, each point representing data for a single site where field data and corresponding digital numbers of Landsat TM bands were collected. Sites are identified as outliers (open boxes) if all 15 sites in the class are strongly separated from the other sites in the data set (●). Outlier classes are indicated by a dashed line. Biplots of variables are shown as arrows from the ordination centroid. (*a*) PCA of 17 classes, outlier class on the left of the diagram is Water, outlier class on the right is Unvegetated Ridge (axis I = 81.05%, axis II = 12.14%). (*b*) PCA of 15 classes (Water and Unvegetated Ridge removed), outlier class is Unvegetated Shoreline (axis I = 59.02%, axis II = 21.83%). (*c*) PCA of 14 classes (Unvegetated Shoreline removed), outlier class is Recent Burn (axis I = 72.06%, axis II = 19.29%). (*d*) PCA of 13 classes (Recent Burn removed), no outlier classes are present (axis I = 72.81%, axis II = 18.59%).

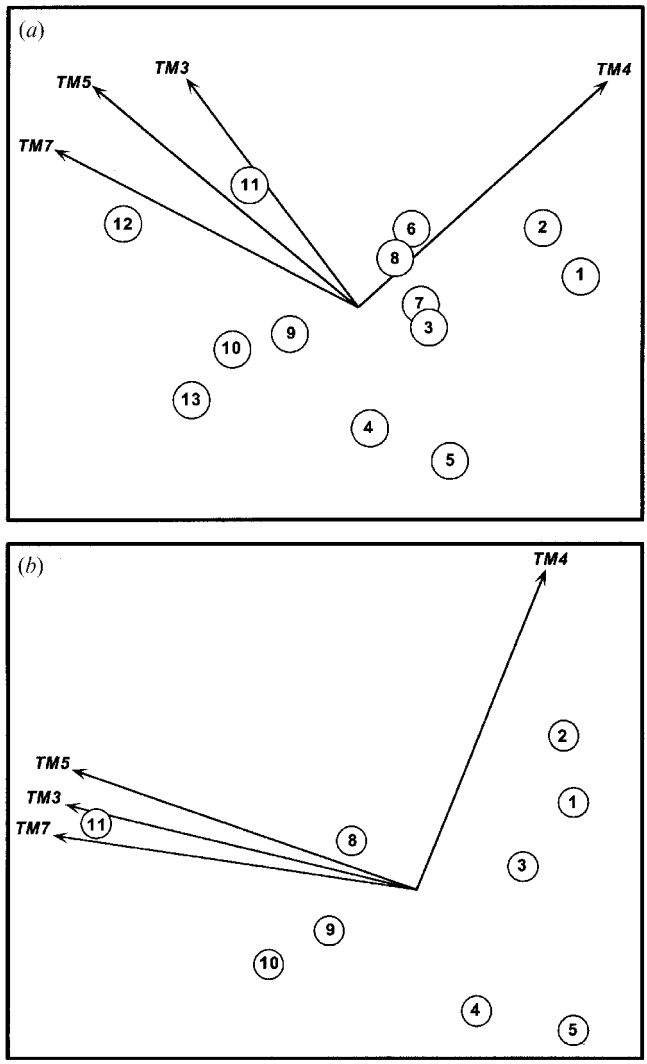

Figure 4. Multiple discriminant analysis of spectral reflectance variables of vegetation classes. 95% confidence ellipses for means are shown; class labels as in table 1. Biplots of variables are shown as arrows from the ordination centroid. (*a*) MDA of 13 vegetation classes (Wilk's $\Lambda, = 0.0010$, $p < 0.001$). (*b*) MDA of nine vegetation classes with classes 6, 7, 12, 13 removed (Wilk's $\Lambda, = 0.0002$, $p < 0.001$).

Each step in the iterative process summarized above revealed various features of the spectral data. As outlier unvegetated groups were removed, band 4 (near infrared) became increasingly important in discerning the vegetated classes. Bands 1–3 (visible) were proximate in all cases, indicating a high degree of multicollinearity. Bands 5 and 7 are also correlated with the visible bands, but band 4 contains unique information not carried by the other bands.

4.2. *Multiple discriminant analysis*

An MDA of the 13 remaining classes revealed varying degrees of interclass separation (figure 4(*a*)). Four of the classes were subsequently masked or combined

with other classes. Classes 3 (Willow Birch Shrub Fen) and 7 (Willow Sedge Poor Fen) show considerable overlap in spectral space. Since they cannot be separated spectrally and are floristically similar, they were combined to form a single class.

Classes 6 (Graminoid Willow Salt Marsh) and 8 (Lichen Spruce Bog) also show considerable overlap in spectral space, but are floristically very different and spatially distinct (Salt Marsh occurs exclusively along the coast, while Lichen Spruce Bog is found farther inland). To distinguish the salt marsh class on the vegetation map, a 15 km wide strip along the shore of Hudson Bay was digitized and masked. Unvegetated areas of the mask were subtracted and an unsupervised 28-class isodata classification was run to spectrally separate salt marsh from other vegetation classes.

Class 12 (*Dryas* Heath) occurs on long, narrow former beach ridges that parallel the Hudson Bay coastline. While limited in spatial extent, this vegetation class represents a distinct and ecologically/culturally significant landscape feature. To ensure that these small, elongated regions were classified with a high level of certainty on the final map, a mask was created for this spectrally distinct class using the same procedure applied to regenerating burns. Pixels were then only identified as being *Dryas* Heath if they occurred under the mask to ensure that no pixels were misclassified as *Dryas* Heath in areas where it is known to not occur.

Class 13 (Regenerating Burn) proved to be spectrally distinct, but it represents a highly variable disturbance feature that spans several vegetation classes. Regenerating burn boundaries were therefore hand-digitized to create a bitmap, and an unsupervised 28-class isodata classification was run under the mask. All classes falling within the spectral range of the ground-sampled regenerating burns were aggregated into a single regenerating burn class, converted to a bitmap mask, and removed from further classification.

A residual MDA (excluding classes 6, 7, 12 and 13) revealed strong separation of the nine remaining classes (figure 4(*b*)). The lichen peat plateau (class 11) has the highest reflectance in bands 3, 5 and 7. The most productive sites (classes 1–3) have highest reflectance in band 4, but low reflectance in bands 3, 5 and 7.

4.3. *Redundancy analysis*

Redundancy analysis indicated that 43% of variance in the CA ordination space can be explained by variation in spectral reflectance (bands 3, 4, 5 and 7). This suggests a reasonably strong relationship between floristic composition and Landsat spectral reflectance, but not surprisingly a large amount of the variation in floristic composition remains unaccounted for. This occurs because spectral reflectance is largely a function of the structural, rather than floristic, properties of vegetation. The RDA ordination (figure 5) indicates that vegetation classes dominated by lichen (classes 8, 10 and 11) are characterized by high reflectance in bands 3, 5 and 7, but low reflectance in band 4. More productive wet fen and shrub habitats (classes 1, 2 and 3) show highest reflectance in band 4, but low reflectance in bands 3, 5 and 7.

5. Discussion

Coarse-scale mapping projects have often downplayed the importance of collecting ground cover data. Indeed, such data are often only used to label classes derived from an unsupervised classification of spectral information (Cihlar 2000). For finer-scale vegetation mapping projects (1:250 000 and finer), our results indicate that detailed and extensive field data are essential to the interpretation and analysis of satellite imagery. We strongly suggest that image classification (whether using a

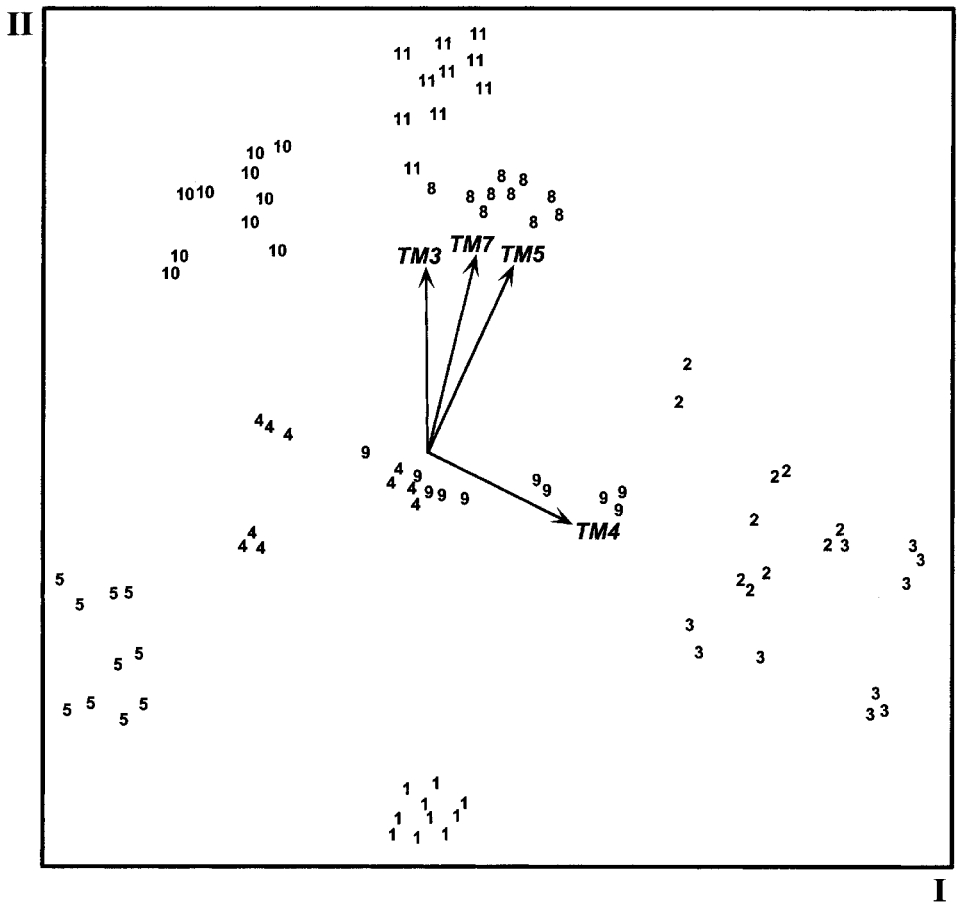

Figure 5. Redundancy analysis relating nine vegetation classes (numbers, see table 1 for codes) and four spectral TM bands (3–5 and 7, displayed as biplots). The two ordination axes I and II account for 77.7% of the vegetation–spectral band relation. Total redundancy is 43%.

supervised or unsupervised approach) should occur only after vegetation classes have been identified and characterized using appropriate statistical analyses of field data. Once this has been done, the spectral information associated with each vegetation class must be examined to ensure that the classes are spectrally distinct. The mapping of ecologically meaningful vegetation classes must begin with a classification of ground vegetation (e.g. Thompson *et al.* 1980, Nilsen *et al.* 1999), not a classification of satellite spectral reflectance data.

The use of ground data to evaluate spectral information provides critical insights into optimal band combinations for classification purposes, and identifies the degree of separability of different classes. It is generally recognized that not all Landsat TM bands are required for classification purposes (Richards 1993). Band selection should be based on the ability of different band combinations to resolve the classes of interest and on the analysis of correlations among bands.

In our initial analyses, PCA was performed using six spectral bands (bands 1–5 and 7) as variables. Including the three highly collinear visible bands 1–3 places

strong emphasis on distinguishing vegetated from unvegetated classes. Band 1 is considered particularly effective in separating vegetation from barren ground (Lillesand and Kiefer 1994). After the unvegetated outliers had been recognized and removed, bands 1 and 2 were excluded from subsequent multivariate analyses (MDA, RDA) in order to reduce variable multicollinearity. When discriminating vegetation types, it is generally recommended that at least one band from each of the visible, near-infrared and mid-infrared regions are included (Beaubien 1994). The combination of bands 3 (visible red), 4 (near infrared) and 5 (mid-infrared) has been widely used (e.g. Benson and Degloria 1985, Horler and Ahern 1986, Moore and Bauer 1990, Beaubien 1994), although band 7 (mid-infrared) is sometimes substituted for band 5. Multivariate biplots (Gabriel 1971) effectively summarize the correlations among bands and trends in the spectral reflectance properties of vegetation classes, thus allowing the analyst to select bands that are most appropriate for a particular application.

The initial 17 land cover/vegetation classes (table 1) displayed some overlap in spectral space. A classification based directly on these 17 groups would therefore have resulted in an inaccurate and misleading vegetation map. In this study, the reasons for overlapping spectral signatures varied, but were largely related to floristic composition, physiognomic structure and substrate composition. For example, classes 3 and 7 (Willow Birch Shrub Fen and Willow Sedge Poor Fen) were floristically similar, and are distinguished mainly by higher shrub cover in the Willow Birch Shrub class. These two classes were amalgamated since this relatively subtle difference in vegetation could not be resolved by the TM spectral reflectance data. In contrast, the spectral overlap between classes 6 and 8 (Graminoid Willow Salt Marsh and Lichen Spruce Bog) illustrates how very different vegetation classes can have nearly identical spectral signatures. The salt marsh class is characterized by small patches ($1–50\,m^2$) of halophytic vegetation (graminoids and low shrubs) interspersed with highly reflective unvegetated clay, while Lichen Spruce bogs are completely vegetated, characterized by a sparse cover of coniferous trees, and a highly reflective ground layer dominated by lichens. These two classes are structurally somewhat similar (patches of highly absorptive vegetation on a highly reflective matrix), but they share no species in common and are functionally very different. They cannot be distinguished using spectral reflectance information alone, and so must be mapped using ancillary information.

6. Conclusion

When mapping vegetation, it is important to recognize that difficulties in characterizing thematic classes are not merely related to a mixed pixel problem. Indeed, very different combinations of vegetation cover and substrate conditions can sometimes produce nearly identical reflectance signatures. This phenomenon can only be identified through extensive collection of field data and a careful examination of its spectral properties. A number of approaches are available to separate distinct vegetation classes showing overlap in spectral space. It is difficult to recommend a single approach, however, since each mapping exercise presents unique problems and challenges. Ancillary information from digital elevation models, aerial photographs, substrate data and so forth have long been used to improve the separation of classes in vegetation mapping exercises (e.g. Nilsen *et al.* 1999, Walker 1999). Alternatively, multi-temporal satellite imagery can be used to separate classes that vary spectrally through time, for example over a growing season (Fuller and Parsell 1990, Lunetta

and Balogh 1999). Whatever method is used, it is of course necessary first to identify and characterize the overlapping classes.

We feel that the production of an effective and optimized vegetation map is best achieved using an adaptive learning process that involves careful examination of the relationship between the ground cover and spectral data. The approach advocated here is labour intensive and involves numerous iterative steps. However, we believe that it provides a highly robust and flexible approach to thematic mapping, since the analyst is able to make informed choices at all stages of the decision-making process. The graphical approach that we employ (figures 3–5) provides an uncomplicated and intuitive method for displaying very complex multivariate relationships between ground cover and spectral reflectance data. It has been our experience that presenting multivariate analysis results only in tabular form, while informative, is rarely enlightening.

7. Summary

A model is a simplification of a natural system that retains the important or usable information while removing 'noise' that would otherwise obscure important trends. The development of landscape-scale vegetation/land cover maps requires a rigorous modelling approach that includes both validation and verification. A graphically-oriented multivariate approach can assist the modelling process by allowing the analyst to consider multiple satellite bands simultaneously, and to identify overlap in spectral reflectance values prior to classification. This is a critical process, since mutually exclusive classes must be defined in order to unambiguously classify pixels in a satellite image.

Acknowledgments

The authors are grateful for support from Wapusk National Park, the Manitoba Department of Conservation, Churchill Northern Studies Centre, the Northern Studies Training Program of the Department of Indian and Northern Affairs, the Canadian Department of Fisheries and Oceans, the Western Canada Service Centre of Parks Canada, the Canadian Wildlife Federation, the Manitoba Chapter of the Wildlife Society, Wat'chee Lodge, the Natural Resources Institute (R. K. Brook) and a Natural Sciences and Engineering Research Council of Canada individual operating grant (N. C. Kenkel). We are grateful to D. Clark for support through all stages of this project. S. McLachlan and T. Naughten provided computer facilities and practical input. Special thanks to E. Richardson for assistance in the field and the Wapusk National Park Warden Service for logistical support. The comments from two anonymous reviewers significantly improved the quality of this paper.

References

ARAI, K., 1992, A supervised Thematic Mapper classification with a purification of training samples. *International Journal of Remote Sensing*, **13**, 2039–2049.

BEAUBIEN, J., 1994, Landsat TM satellite images of forests: from enhancement to classification. *Canadian Journal of Remote Sensing*, **20**, 17–26.

BENSON, A. S., and DEGLORIA, S. D., 1985, Interpretation of Landsat-4 Thematic Mapper and Multispectral Scanner Data for forest surveys. *Photogrammetric Engineering and Remote Sensing*, **51**, 1281–1289.

BLISS, L. C., COURTIN, G. M., PATTIE, D. L., RIEWE, R. R., WHITFIELD, D. W. A., and WIDDEN, P., 1973, Arctic tundra ecosystems. *Annual Review of Ecology and Systematics*, **4**, 359–399.

CAMPBELL, M., 1995, *The winter ecology of the Cape Churchill caribou* (Rangifer tarandus *ssp.*). M.Sc. Thesis, University of Manitoba, Winnipeg, Manitoba, Canada.

CIHLAR, J., 2000, Land cover mapping of large areas from satellites: status and research priorities. *International Journal of Remote Sensing*, **21**, 1093–1114.

CIHLAR, J., BEAUBIEN, J., XIAO, Q., CHEN, J., and LI, Z., 1997, Land cover of the BOREAS Region from AVHRR and Landsat data. *Canadian Journal of Remote Sensing*, **23**, 163–175.

FERGUSON, R. S., 1991, Detection and classification of Muskox habitat on Banks Island, Northwest Territories, Canada, using Landsat Thematic Mapper Data. *Arctic*, **44 (Supp. 1)**, 66–74.

FOODY, G. M., 1999, The continuum of classification fuzziness in thematic mapping. *Photogrammetric Engineering and Remote Sensing*, **65**, 443–451.

FULLER, R. M., and PARSELL, R. J., 1990, Classification of TM imagery in the study of land use in lowland Britain: practical considerations for operational use. *International Journal of Remote Sensing*, **11**, 1901–1917.

GABRIEL, K. R., 1971, The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, **58**, 453–467.

GANTER, B., COOKE, F., and MINEAU, P., 1996, Long-term vegetation changes in a Snow Goose nesting habitat. *Canadian Journal of Zoology*, **74**, 965–969.

GITTINS, R., 1985, *Canonical Analysis* (Berlin: Springer).

GREEN, R. H., 1993, Relating two sets of variables in environmental studies. In *Multivariate Environmental Statistics*, edited by G. P. Patil and C. R. Rao (Amsterdam: Elsevier), pp. 149–163.

HOMER, C. G., RAMSEY, R. D., EDWARDS, T. C. JR, and FALCONER, A., 1997, Landscape cover-type modelling using a multi-scene Thematic Mapper mosaic. *Photogrammetric Engineering and Remote Sensing*, **63**, 59–67.

HOPE, A. S., KIMBALL, J. S., and STOW, D. A., 1993, The relationship between tussock tundra spectral reflectance properties and biomass and vegetation composition. *International Journal of Remote Sensing*, **14**, 1861–1874.

HORLER, D. N. H., and AHERN, F. J., 1986, Forestry information content of Thematic Mapper data. *International Journal of Remote Sensing*, **7**, 405–428.

HORN, L. N. G., 1981, *Vegetation mapping in northern Manitoba with Landsat: preliminary assessment of barren-ground caribou wintering range*. Master of Natural Resource Management Practicum, University of Manitoba, Winnipeg, Manitoba, Canada.

JEFFERS, J. N. R., 1982, *Modelling* (London: Chapman and Hall).

KOROBOV, R. M., and RAILYAN, V. Y., 1993, Canonical correlation relationships among spectral and phytometric variables for twenty winter wheat fields. *Remote Sensing of Environment*, **43**, 1–10.

LEGENDRE, P., and LEGENDRE, L., 1998, *Numerical Ecology*, 2nd edition (Amsterdam: Elsevier).

LILLESAND, T. M., and KIEFER, R. W., 1994, *Remote Sensing and Image Interpretation*, 3rd edition (New York: Wiley).

LUNETTA, R. S., and BALOGH, M. E., 1999, Application of multi-temporal Landsat 5 TM imagery for wetland identification. *Photogrammetric Engineering and Remote Sensing*, **65**, 1303–1310.

MATTHEWS, E., 1983, Global vegetation and land use: new high-resolution data bases for climate studies. *Journal of Climate and Applied Meteorology*, **22**, 474–487.

MATTHEWS, S. B., 1991, An assessment of bison habitat in the Mills/Mink Lakes Area, Northwest Territories, using Landsat Thematic Mapper data. *Arctic*, **44 (Supp. 1)**, 75–80.

MATVEYEVA, V., 1994, Floristic classification and ecology of tundra vegetation of the Taymr peninsula, northern Siberia. *Journal of Vegetation Science*, **5**, 813–828.

MOORE, M. M., and BAUER, M. E., 1990, Classification of forest vegetation in north-central Minnesota using Landsat Multispectral Scanner and Thematic Mapper data. *Forest Science*, **36**, 330–342.

MORRISON, R. I. G., 1997, The use of remote sensing to evaluate shorebird habitats and populations on Prince Charles Island, Foxe Basin, Canada. *Arctic*, **50**, 55–75.

MULLER, S. V., RACOVITEANU, A. E., and WALKER, D. A., 1999, Landsat MSS-derived land-cover map of northern Alaska: extrapolation methods and a comparison with

photo-interpreted and AVHRR-derived maps. *International Journal of Remote Sensing*, **20**, 2921–2946.

NEMANI, R., and RUNNING, S., 1997, Land cover characterization using multitemporal red, near-IR, and thermal-IR data from NOAA/AVHRR. *Ecological Applications*, **7**, 79–90.

NILSEN, L., ELVEBAKK, A., BROSSARD, T., and JOLY, D., 1999, Mapping and analysing arctic vegetation: evaluating a method coupling numerical classification of vegetation data with SPOT satellite data in a probability model. *International Journal of Remote Sensing*, **20**, 2947–2977.

O'NEILL, R. V., MILNE, B. T., TURNER, M. G., and GARDNER, R. H., 1988, Resource utilization scales and landscape pattern. *Landscape Ecology*, **2**, 63–69.

PALA, S., and BOISSONNEAU, A., 1982, Wetland classification maps of the Hudson Bay Lowland. *Le Naturaliste Canadien*, **109**, 653–659.

PALA, S., and WEISCHET, W., 1982, Toward a physiographic analysis of the Hudson Bay–James Bay Lowland. *Le Naturaliste Canadien*, **109**, 637–651.

PCI GEOMATICS, 1998, *ImageWorks Version 6.3 EASI/PACE*. Richmond Hill, Ontario.

PEARCE, C. M., 1991, Mapping muskox habitat in the Canadian High Arctic with SPOT satellite data. *Arctic*, **44 (Supp. 1)**, 49–57.

PERRY, C. R. JR, and LAUTENSCHLAGER, L. F., 1984, Functional equivalence of spectral vegetation indices. *Remote Sensing of Environment*, **14**, 169–182.

PETZOLD, D. E., and GOWARD, S. N., 1988, Reflectance spectra of subarctic lichens. *Remote Sensing of Environment*, **24**, 481–492.

PODANI, J., 1994, *Multivariate Data Analysis in Ecology and Systematics* (The Hague: SPB Academic Publishing).

RICHARDS, J. A., 1993, *Remote Sensing Digital Image Analysis: An Introduction*, 2nd edition (New York: Springer).

RICHARDSON, A. J., and WIEGAND, C. L., 1977, Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing*, **43**, 1541–1552.

RITCHIE, J. C., 1956, The native plants of Churchill, Manitoba. *Canadian Journal of Botany*, **34**, 269–320.

RITCHIE, J. C., 1962, *A geobotanical survey of northern Manitoba*. Arctic Institute of North America Technical Paper No. 9.

ROUGHGARDEN, J., RUNNING, S. W., and MATSON, P. A., 1991, What does remote sensing do for ecology? *Ecology*, **72**, 1918–1922.

SHASBY, M., and CARNEGGIE, D., 1986, Vegetation and terrain mapping in Alaska using Landsat MSS and digital terrain data. *Photogrammetric Engineering and Remote Sensing*, **52**, 779–786.

SIMS, R. A., COWELL, D. W., and WICKWARE, G. M., 1982, Classification of fens near southern James Bay, Ontario, using vegetational physiognomy. *Canadian Journal of Botany*, **60**, 2608–2623.

THOMPSON, D. C., KLASSEN, G. H., and CIHLAR, J., 1980, Caribou habitat mapping in the southern District of Keewatin, NWT: An application of digital Landsat data. *Journal of Applied Ecology*, **17**, 125–138.

TOWNSHEND, J. R. G., 1992, Land cover. *International Journal of Remote Sensing*, **13**, 1319–1328.

VOGELMANN, J. E., and MOSS, D. M., 1993, Spectral reflectance measurements in the Genus *Sphagnum*. *Remote Sensing of Environment*, **45**, 273–279.

WALKER, D. A., 1999, An integrated vegetation mapping approach for northern Alaska (1:4 M scale). *International Journal of Remote Sensing*, **20**, 2895–2920.

WEBBER, P. J., RICHARDSON, J. W., and ANDREWS, J. T., 1970, Post-glacial uplift and substrate age at Cape Henrietta-Maria, southeastern Hudson Bay, Canada. *Canadian Journal of Earth Sciences*, **7**, 317–325.