INT. J. REMOTE SENSING, 1999, VOL. 20, NO. 17, 3311–3331

Evaluation of MK-4 multispectral satellite photography in land cover classification of eastern Ontario

D. J. KING,† M. JOLLINEAU‡ and B. FRASER§

†Department of Geography, Carleton University 1125 Colonel By Drive, Ottawa, Ontario, Canada K1S 5B6
‡Department of Geography, Faculty of Environmental Studies, University of Waterloo, Waterloo, Ontario, Canada
§Parks Canada, Ontario Region, Cornwall, Ontario, Canada

(*Received 15 April 1997; in final form 18 June 1998*)

Abstract. Russian MK-4 multispectral satellite photography has been investigated for potential in land cover classification. Thematic maps were generated using maximum likelihood, neural network and context classifiers. Classifications of the raw spectral data, of spectral transforms, and of combined spectral/textural data were evaluated. Low point-based class accuracies resulted for land cover types exhibiting high spatial variability at the given pixel spacing of 7.5 m, while more spatially homogeneous cover types were well classified. Several issues arose which need to be addressed for effective future use of high-resolution satellite sensors in regional land cover mapping. They include the need for further research in techniques for classification and accuracy assessment which are sensitive to the spatial variance of such high resolution imagery, and optimization of class attribute definitions.

1. Introduction

As decision making for resource and environmental management becomes more complex, the need for more precise spatial information with greater positional and thematic accuracy increases. In regional land management, there is currently a need for greater spatial resolution in small scale digital maps covering large areas. Stable, accurate sources of land cover information are required for decision making on a temporal basis. In eastern Ontario, identification of patches smaller than those which can be accurately mapped with Landsat TM (Hansen 1994) or SPOT is a necessity as they contribute significantly to the ecology of the highly fragmented landscape.

The research presented in this paper was conducted within an operational framework: Parks Canada, a section of the Heritage Canada Ministry, had identified specific mapping needs and desired to investigate alternative data sources to Landsat TM and SPOT, which have been found not to be suitable for environmental management of a wide variety of land cover units. The selected data source, MK-4 multispectral photography, was assessed using both conventional classification techniques to provide a base for comparison with other data types, and more experimental

e-mail: dking@ccs.carleton.ca

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online
http://www.tandf.co.uk/JNLS/res.htm
http://www.taylorandfrancis.com/JNLS/res.htm

techniques and data transforms. MK-4 multispectral photography is acquired aboard the Russian Resours-F satellite. It consists of three spectral bands in the green (515–565 nm), red (635–690 nm), and near-infrared (810–900 nm) with coverage of approximately 165 × 165 km and 7.5 m nominal ground pixel spacing. The spectral bands are positioned at the same wavelengths as bands 1–3 of Landsat TM and SPOT, and the respective bandwidths are smaller. At a lower cost than Landsat or SPOT imagery, these MK-4 imagery specifications indicated a strong potential as an alternative data type for regional land cover mapping.

The ground pixel size of 7.5 m is much smaller than Landsat TM (30 m) and SPOT (20 m) so the spatial resolution of land cover classes covering small areas and of within-class detail was expected to be much greater. In particular, in landscapes such as southern Ontario, which are highly fragmented from long periods of clearing and land division into a mosaic of farms and woodlots, there is a need to map smaller patch sizes than is possible with lower resolution satellite imagery. Assessment of the MK-4 data was also considered to be important as a preliminary indication of the capabilities and issues expected to arise when anticipated 1–5 m high-resolution multispectral satellite imagery becomes available for use in regional thematic mapping.

Associated with the potential for higher detail mapping and for mapping of smaller areas is, however, an expected increase in within-class variance which may prove disadvantageous to conventional per pixel classifiers. For example, it has been often reported that pixel sizes larger than 30 m produce better accuracy of standard classes such as USGS Level I and II (Anderson *et al.* 1976) using per-pixel classification and point-based accuracy assessment (e.g. Moody and Woodcock 1994, Martin *et al.* 1988). Consequently, in an attempt to capture the high spatial information content of the MK-4 data, this research included evaluation of both a spatially sensitive context classifier and texture transforms.

1.1. *Research objectives*

The primary objective was to evaluate MK-4 multispectral satellite imagery in land cover classification using both standard and more experimental image analysis techniques. Included as standard techniques were per pixel unsupervised clustering and maximum likelihood classification of raw data, band ratios, principal components, and noise reduced data. They were selected to provide classifications which could be compared with those typically obtained from SPOT or Landsat data and to provide references for evaluation of the more experimental techniques. The experimental techniques included addition of image texture transforms in classification, and classification using neural network and context classifiers.

The research components of this paper are thus twofold: first, a new high-resolution data source is investigated for regional thematic mapping; and second, given the high spatial resolution and apparent texture in the imagery, alternative classifiers are considered and spatial information is incorporated into the classification process.

2. Image data and attribute scheme

2.1 *MK-4 image data*

The MK-4 image data of eastern Ontario, Canada and upper New York State, USA was acquired on 21 May, 1991. This was the best imaging date available to Parks Canada at the time of initiation of the research in 1994. The data consisted of three aligned photographs taken simultaneously with different cameras, each equipped with a spectral bandpass filter as listed above. The coverage of each photograph was approximately 165× 165 km as shown in figure 1.

The photographs had been scanned in Moscow at 2988 dpi (8.5 μm) producing images of 21 883× 21 883 pixels, each pixel being eight-bit and representing 7.5× 7.5 m on the ground. They were then geo-referenced to a NAD27 datum UTM grid and tiled into six scenes of 10 940 columns by 7300 or 7280 rows by a contractor in the USA. For the purposes of this study, the scenes were numbered 1–6. Scenes 1 and 2 were in the NW and NE corners with Scene 2 extending into the southern part of the city of Ottawa. Scenes 3 and 4 were in the middle west and east portions, with a large part of Scene 4 being in north-western New York. Scenes 5 and 6 were in the south-west and south-east corners with Scene 6 being located entirely in New York.

A sample of the three raw spectral bands from Scene 2 is shown at full resolution in figure 2. The Ottawa, Ontario airport is in the centre of the images. Scene 2 was used for most of the detailed analysis of classification input parameters, data transformations, and additional classifiers.

2.2. *Attribute definition for classification*

Given the context of the research as outlined above, the design of an attribute scheme had to consider the following criteria: (a) be representative of the classes which are needed for decision making, and (b) be representative of the expected high spatial and spectral variation of the data. The attribute scheme listed below was decided on by balancing these two principal requirements. The classes were:

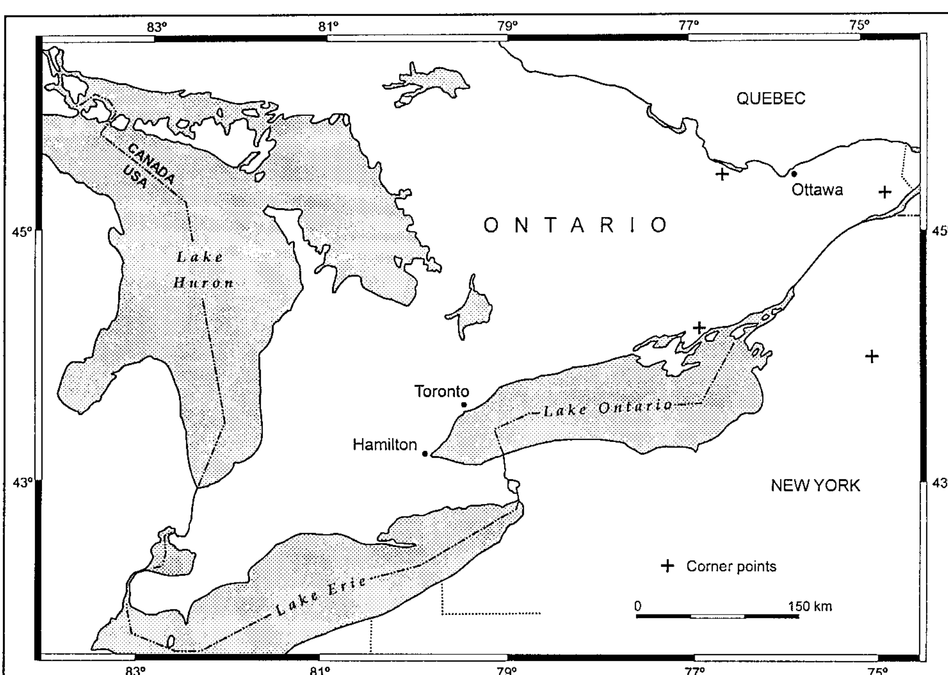
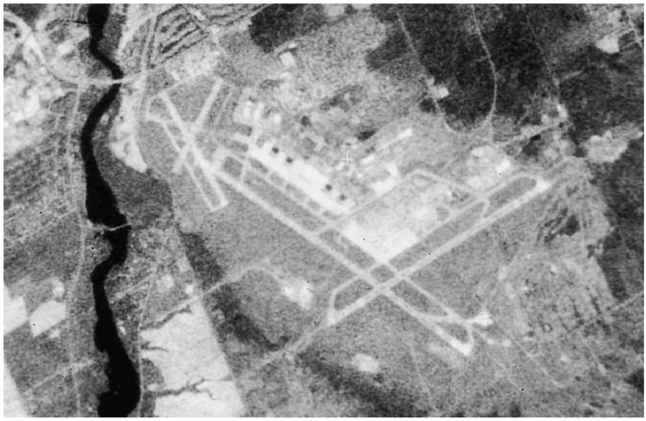


Figure 1. Eastern Ontario and upper New York State with the image corners marked by the symbol '+'.



(*a*)

(*b*)

(*c*)

Figure 2. Portion of raw Scene 2 MK-4 data displayed at full resolution. Spectral bands are: (*a*) green (515–565 nm), (*b*) red (635–690 nm), and (*c*) near-infrared (810–900 nm).

1. Urban impervious and bare rock.
2. Urban grass, pasture and new crops.
3. Bare soil.
4. Forest: coniferous closed canopy.
5. Forest: coniferous open canopy.
6. Forest: mixed coniferous and deciduous.
7. Forest: deciduous closed canopy.
8. Forest: deciduous open canopy.
9. Early successional: shrubs, sparse trees, mostly open.
10. Deep water.
11. Shallow water or sediment loaded water.
12. Wetland: swamp.
13. Wetland: marsh.
14. Wetland: fen.
15. Wetland: bog.

Areas of any of these classes as small as 15 m in diameter were required (about 2×2 pixels). This was used as a minimum mapping unit in field verification and accuracy assessment.

The urban classes were included for identification of development which had taken place on Parks Canada lands. Classes 4–8 were selected to represent the deciduous, coniferous and mixed forest classes typically used in classification of raw Landsat TM data at USGS Level II but with greater detail representative of the potential resolution of canopy structure. Closed canopies were assumed to represent a more mature and perhaps later successional stage than open canopies which would be more influenced by intermediate and shrub species. A division of 50% cover was made to define ‘open’ and ‘closed’ canopies. For the coniferous forest classes (4, 5), large natural stands were expected to be relatively rare. Instead, plantations consisting mostly of pines (Scots pine—*Pinus sylvestris*, red pine—*Pinus resinosa*, jack pine—*Pinus banksiana*, and white pine—*Pinus strobus*), spruce (generally white spruce—*Picea glauca*) and some larch (*Larix laricina*) comprise most of the coniferous forests. The ‘open’ and ‘closed’ canopy designations therefore divided young and mature plantations in most cases. The two major deciduous groups in the region were expected to be usually assigned to one of the two classes. Open canopy forests are dominated by pioneer species such as birch (mostly white birch—*Betula papyrifera*) and poplar (mostly aspen—*Populus tremuloides* and *Populus grandidentata*), which rarely reach levels of canopy closure above 50%, except at very old age during transition to a later successional stage and species composition. The other group, dominated by nonpioneer sugar maple (*Acer saccharum*), beech (*Fagus grandifolia*), and ash (mostly white ash—*Fraxinus americana*), usually has a high degree of canopy closure and often succeeds the pioneer forests. Class 9 represented a transitional class typically occurring in abandoned agricultural fields. For the water classes, a distinction was made between shallow and deep water depending on whether emergent vegetation or the bottom of the water body were visible. Classes 12–15 were the major wetland groups important to environmental management. For the purposes of efficient field work, swamp (class 12) was defined as forested wetland with willow (*Salix spp.*) species present and some dead or dying trees due to a high water table. Marsh (class 13) was defined as nonforested wetland with cattails (*Typha spp.*) and other tall grasses. Fen (class 14) was defined as having standing water which may be running, with shrubs and hummocks of grassy species. Bog (class 15) was defined as a depression or raised encroachment deposit with sphagnum moss present. Fens and bogs are rare in this region and difficult to identify in the field.

3. Methods

3.1. *Unsupervised spectral cluster analysis*

Unsupervised spectral cluster analysis was initially conducted on Scene 3 because it contained a balanced mix of agricultural terrain and the rugged, less developed Canadian Shield (i.e. it consisted of representative samples of all the terrain types present in the other scenes). The potential number of spectrally distinct clusters in the data was determined using the Narendra–Goldberg clustering algorithm (Narendra and Goldberg 1977). This algorithm performs a nonparametric multi-dimensional histogram analysis by dividing the three dimensional histogram of sample data into uni-modal histogram clusters.

A second clustering procedure was conducted to produce a cluster map for use in guiding training site selection. It used a parametric *k*-means clustering algorithm commonly called ISODATA (Tou and Gonzalez 1977). The number of distinct clusters determined by the Narendra–Goldberg techniques was to have been used in the *k*-means procedure as the maximum number of clusters to determine, while the minimum number of clusters was selected as 16 (the 15 defined classes plus one representing areas outside scene boundaries which were artefacts of image registration). Clusters were merged if their separation in feature space was below a threshold of one. If the standard deviation of a cluster was greater than 10, the cluster was split. A maximum number of iterations of 20 was specified, although 15 were sufficient to reduce movement of all cluster means in feature space to less than 1% of the mean values.

The cluster maps were visually analysed to determine: (a) suitable sites for training and test site selection in clusters which could be matched to one of the desired classes, and (b) if any clusters which could not be identified could be visited to determine if they were a different class from those of the defined class list, or if they represented a spectral variation of one of the desired classes.

3.2. *Training and test site selection*

In field data acquisition during May–July 1995, routes of approximately 600 km were designed to traverse the complete coverage of each scene. At least four sites were visited for each class in each scene. About a third of the sites were selected with the aid of the cluster maps and 1∶10 000 air photos and maps. The remaining sites were selected by chance encounter en route. An effort was made to select sites of a given class far apart from each other to avoid spatial autocorrelation and to represent the natural site variability of each class in the region. The minimum number of sites to be visited and analysed was: 4 sites $\times$ 15 classes $\times$ 6 scenes $=$ 360. Sample size was varied from many pixels to small samples on the order of the 2×2 pixel minimum mapping unit in order to account for the wide variety of patch sizes of each class. The selected sites were divided equally between training and test samples. At each site current site class and conditions were noted. If it could be determined or was known that the site conditions would have been different in May 1991, this information was also noted.

3.3. *Spectral data analysis and classification*

For each training class the mean, standard deviation, variance and histogram in each band were analysed to ensure normality and eliminate outliers. Additional analysis included plotting of red *versus* near-infrared training data (scatterplots), and separability analysis using the J–M distance (Jensen 1996).

Classification was initially conducted on all scenes using a maximum likelihood classifier. The *a priori* probabilities were assumed to be equal, as mapping at this spatial resolution and attribute detail had not been previously conducted. This also avoided the bias of varying these proportions subjectively. To determine the potential contribution of a null class and the cut-off probability threshold to classification accuracy, three test classifications of Scene 2 were conducted: (1) null class included, threshold = 3 standard deviations (virtually all training data assumed to be representative of a given class); (2) null class not included, threshold = 3; and (3) null class included, threshold = 2 (95% cut-off bounds for training data). Scene 2 was also used to determine if certain spectral transformations, textural transformations, or other classifiers could improve the classification accuracy. All resulting thematic maps were filtered using a 3×3 mode filter.

3.3.1. *Evaluation of transformed data*

The Scene 2 data were modified or transformed using the following techniques in attempts to improve classification accuracy. *Principal components analysis* (PCA) was conducted with the expectation that: (a) the first two components would be used in the classification as the third would be mostly noise; and (b) that processing time would be reduced from the three band case. *Spectral band ratios*—the normalized difference vegetation index, [NDVI = (near-infrared − red) ÷ (near-infrared + red)], and a second difference ratio [NDVI′ = (near-infrared − green) ÷ (near-infrared + green)] were used, the latter selected because a minimum of two channels is required in maximum-likelihood classification. *Texture transformation*—the high spatial resolution of the data was expected to exhibit greater texture variations between land cover types than would typically be present in Landsat or SPOT images. The co-occurrence texture measure, ‘Contrast’ was selected based on previous experience with it (e.g. Yuan *et al.* 1991) and based on visual evaluation of texture images produced by a variety of texture measures (e.g. entropy, angular second moment, standard deviation, etc.). Co-occurrence texture measures utilize a matrix which tabulates the frequency of occurrence of adjacent pixels of all possible grey level pairs. Low texture (a spatially uniform land cover) produces high frequencies near, or on the diagonal of the co-occurrence matrix, while spatially variable land cover types exhibiting high texture will produce high frequencies of pixel pairs farther from the diagonals. The ‘Contrast’ measure sums the frequency in each entry of the co-occurrence matrix times the entry’s squared distance from the diagonal. The following input variable values were specified: (1) sampling direction to the right of any pixel (arbitrary); (2) a sample distance of 1 pixel (adjacent pixels were compared); and (3) a window size of 3×3 to extract texture that would be present in areas close to the minimum mapping unit in size. The resulting texture image, scaled to 0–255, was then used with the three raw spectral bands in a four-band classification.

3.3.2. *Evaluation of context and neural network classifiers*

To determine whether accuracy could be improved through the use of non-parametric classification, two other classifiers were tested on Scene 2. The first, a Context classifier incorporates spatial information by evaluating the frequency distribution histograms of the training data and assigning a centre pixel within a specific window size to the closest class based on the minimum 'city block' distance between means. The algorithm works not with the raw data but with a grey level vector reduction image essentially created by transforming the data into eigenspace. This algorithm is fully described in Gong and Howarth (1992). Additional tests of land cover classification using the algorithm are given in Treitz *et al.* (1992). In this research, two window sizes were tested, 7×7 and 21×21. The former was the minimum which could produce suitable histograms for the classifier. The latter, the maximum allowable, was selected to approximate the physical area in which Gong and Howarth obtained their best results in similar terrain using SPOT data.

The second classifier was a feed forward neural network with back propagation training (Maren *et al.* 1980, Foody and Arora 1997, Kanellopoulos and Wilkinson 1997, Yool 1998). The neural network consisted of an input layer of image channels, an output layer of the training classes, and a single hidden layer with the same number of hidden units as output classes. To avoid subjective manipulation of input parameters, default values were initially used: a learning rate of 0.1, a momentum rate of 0.9, a maximum total error of 0.01, a maximum individual error of 0.001, and a maximum number of iterations of 1000. Upon determining that low class accuracies had been achieved, the learning rate was reduced to 0.05, the momentum rate was reduced to 0.3, and the number of iterations was increased to 2000. The large amount of time required for training and classification did not permit many trial runs with varied input parameters as is common in studies of neural network classification.

3.4. *Classification accuracy assessment*

Each classification was evaluated for accuracy by cross-tabulating the map with the test site polygons. The Producer's accuracy associated with errors of omission, the Consumer's accuracy associated with errors of commission, the average accuracy normalized by class sample size, and the kappa coefficient of agreement (the proportion correct above the accuracy which would have resulted had pixels been assigned randomly to classes) were computed (see Congalton 1991). This accuracy assessment was repeated for a set of aggregated classes to see potential improvements in accuracy using lower precision attributes.

4. Results

Visual evaluation of the imagery revealed several characteristics. (1) The spatial variability of the data was very high. Random variation in grey tone or colour was evident in water bodies and agricultural fields of essentially uniform cover. This variability appeared to be a result of both the small pixel spacing and of the image formation process (photography and/or scanning) as it varied significantly between spectral bands. (2) A blocky pattern resulting from the film scanning process was evident in contrast stretched images of dark features such as water. It consisted of rectangular blocks of varied size and brightness, each having a grey level of between one and four digital numbers different from adjacent blocks. It was expected that such artificial variation in grey levels in each band would affect class separability. (3) Scenes 5 and 6 at the south end of the imagery were almost half covered by cloud or cloud shadow. The clouds were mostly aggregated, blocking the terrain entirely, but there were also several random patches of cloud dispersed throughout both of these scenes.

4.1. *Unsupervised clustering*

The Narendra–Goldberg clustering procedure produced 120 distinct spectral clusters in many very small polygons which exhibited high spatial variance similar to the raw images. This number of clusters was to have been used as the maximum number in the subsequent ISODATA procedure but since it was so large, an arbitrary value of 30 was selected. The results showed that 29 distinct clusters were found as two had been merged. Assigning the clusters to specific classes for use in field site selection was straightforward for larger agricultural, forest and water classes. However, it was much more difficult to select small clusters on the order of the minimum mapping unit and try to find them in the field. Errors in the image co-ordinate system, the 1: 10 000 maps, and the differential GPS used in the field were too great for navigation to pre-selected points. Thus, in field verification, the cluster maps were used as a guide for large site selection while smaller sites were selected from air photos and chance encounters of representative sites while en route.

4.2. *Spectral data analysis*

A typical scatterplot of the red and near-infrared bands in Scene 2 is shown in figure 3. The numbered class means are encompassed by an ellipse of two standard deviations. The scatterplot excludes the water classes (10, 11) because they were very distinct from the rest. There were large overlaps in many of the classes: class 15 (bog) was almost completely contained within class 4 (closed coniferous forest), class 5 (open coniferous) overlapped strongly with many classes (2, 9, 12, 4, 15, 6), classes 7 and 8 (open and closed deciduous forest) overlapped strongly, class 6 (mixed forest) had a small area that was distinct, but most of it overlapped with its two constituents, deciduous and coniferous forest. Other than the water classes, the most spectrally distinct classes were bare soil (3), and early successional (9). The separability matrix for the training data of Scene 2 is shown in table 1.

Normalized values of the J–M distance, with an output range of 0–2 are given. Values in the range of 0–1 for the separation of two classes in spectral space are

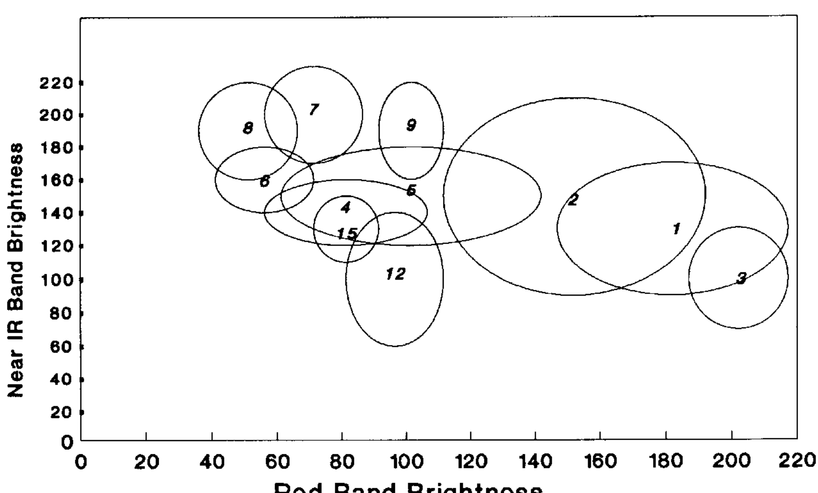


Figure 3. Scatterplot of training data extracted from the red and near-infrared bands for each class in Scene 2.

Table 1. Separability matrix showing normalized (0–2) J–M distance for all training class pairs (listed by class number) and Scene 2 raw data.

	1	2	3	4	5	6	7	8	9	10	11	12
2	1.3754											
3	0.6302	1.7123										
4	1.9927	1.7914	1.9999									
5	1.8863	1.5460	1.8646	1.0438								
6	1.9996	1.9766	2.0000	1.4263	1.4355							
7	1.9997	1.9866	2.0000	1.8566	1.7592	1.2058						
8	1.9998	1.9885	2.0000	1.8651	1.7016	0.8399	0.4951					
9	1.9767	1.6936	1.9990	1.7248	1.4144	1.9196	1.8432	1.9055				
10	1.9999	2.0000	1.9990	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000			
11	1.9996	1.9996	1.9992	1.9997	1.9934	1.9949	1.9999	1.9999	2.0000	1.9626		
12	1.9378	1.4776	1.9758	1.5597	1.6138	1.9769	1.9905	1.9945	1.7781	2.0000	1.9970	
15	1.9999	1.9946	2.0000	1.8587	1.7307	1.4003	1.9791	1.8896	1.9999	2.0000	1.9548	1.9995

considered to be ‘very poor’, values between 1.0 and 1.9 are considered ‘poor’, although classification may result in a map with more information than noise, and a range from 1.9 to 2.0 is considered ‘acceptable’. The results reflect those described above from the scatterplot. Only the water classes were well separated from all other classes and from each other. Of the remaining 59 class pairs which don’t include a water class, four were ‘very poorly’ separated ((7, 8), (1, 3), (6, 8), (4, 5)) and an additional 17 were ‘poorly’ separated.

4.3. *Scene classification/thematic mapping*

The following sections summarise the results of multispectral classification of the raw spectral bands, transformed bands and other classifiers. An example error matrix is given to demonstrate the analysis which was conducted in each case. Aggregated information from these error matrices is also presented.

4.3.1. *Raw data classifications*

Before classifying the raw data, the tests of classification threshold and presence/absence of a null class were conducted. Of the three tests described previously, a threshold of 3 standard deviations (3 SD) with a null class included was selected because the average classification accuracy was slightly higher than for the other tests. In addition, a threshold of 3 SD resulted in 15% of the classified pixels being assigned to the null class *versus* 30% for a threshold of 2 SD. The latter was considered unacceptable. However, a threshold of 2 SD was used for bare soil (3) and early successional (9) because they were the only classes for which both the Producer’s and Consumer’s accuracies were greater (by up to 6%) than when a threshold of 3 SD was used. Table 2 shows an example error matrix for the classification of raw data from Scene 2. Table 3 summarises the raw data classification results for Scenes 1–6.

The standard deviation of the average accuracy for each of the six scenes in table 3 varied from 30 to 40%, indicating a wide variability in class accuracies. In most scenes the total range of accuracy was from close to 0.0% for classes such as urban grass/pasture/new crops (class 2) to close to 100% for deep water (class 10), soil (class 3), deciduous classes (class 7 or 8 in some scenes), and, in some scenes, early successional (class 9).

Table 2. Example error matrix for the maximum likelihood classification of raw Scene 2 data.

	Class	Pixels	Null	Map classification													Producer's accuracy (%)
				1	2	3	4	5	6	7	8	9	10	11	12	15	
Test sites	1	98	9.2	85.7	0	5.1	0	0	0	0	0	0	0	0	0	0	85.7
	2	93	36.6	9.7	1.1	0	0	52.7	0	0	0	0	0	0	0	0	1.1
	3	422	0.7	23.2	13	35.3	0	24.6	0	0	0	0	0	0	3.1	0	35.3
	4	351	5.4	0	0	0	9.1	7.1	63.8	0	0	0	0	0	0	14.5	9.1
	5	150	0	0	0	0	42.7	57.3	0	0	0	0	0	0	0	0	57.3
	6	2088	3.1	0	0	0	4.4	12	53.6	0	2.8	0	0	0	0	24.1	53.6
	7	218	11.9	0	0	0	0	0	0	77.5	10.6	0	0	0	0	0	77.5
	8	731	1.5	0	0	0	0.5	0.5	6.4	62.5	26.3	2.2	0	0	0	0	26.3
	9	93	0	0	5.4	0	4.3	5.4	0	0	0	84.9	0	0	0	0	84.9
	10	25 297	3.1	0	0	0	0	0	0	0	0	0	96.9	0	0	0	96.9
	11	165	73.3	0	0	0	0	0	0	0	0	0	0	26.1	0.6	0	26.1
	12	159	0	0	1.9	0	43.4	20.1	0	0	0	3.8	0	0	30.8	0	30.8
	15	10 227	1.4	0	0	0	54.3	9.4	27.8	0	0	0	0	0	0	7.1	7.1
Consumer's accuracy (%)				72.3	5.1	87.4	5.7	30.3	35.4	55.4	66.2	93.4	100	100	89.3	15.5	

Mean Producer's accuracy: 45.50%; mean Consumer's accuracy: 58.20%; kappa: 0.42.

Table 3. Summary of average Producer's accuracy, Consumer's accuracy and kappa coefficient of agreement for each scene derived from the raw data.

Scene	Average Producer's accuracy (%)	Average Consumer's accuracy (%)	Kappa
1	50.4	55.3	0.47
2	45.5	58.2	0.42
3	44.1	48.1	0.40
4	59.2	60.2	0.56
5	34.1	43.8	0.30
6	28.8	47.5	0.27

Tables 4 and 5 list those classes which equalled or exceeded 50% and 75% Producer's or Consumer's accuracy, respectively in the given scenes. The scene numbers are the table entries.

These results show that only the bare soil, deep water and perhaps early successional classes can be consistently mapped using supervised per-pixel maximum-likelihood classification of the three raw spectral bands. Some scene maps were individually suitable for other classes. Training and test samples were not found for all wetland classes in each scene. In particular, bogs and fens were rare and difficult to find en route (one or two typically per scene).

Scenes 5 and 6, in the southern portion of the dataset, produced the lowest classification accuracies. Both were covered by a large amount of visible cloud. Although training and test sites were collected outside the visible cloud regions, it is suspected that the slight fuzziness in the imagery was a consequence of thin cloud, nonvisible to the eye but significant enough to impact the band statistics.

In addition to variability between scenes produced by clouds, a comparison of mean values of training samples from the same classes in Scenes 3 and 4 (adjacent to each other in the middle of the photograph) revealed that the brightness of these

Table 4. Scene numbers in which each class equalled or exceeded 50% Producer's or Consumer's accuracy in raw data classification.

Class	Producer's accuracy> 50%	Consumer's accuracy> 50%
1. Impervious	2, 4	4, 5
2. Pervious		1
3. Soil	1, 4, 5	1, 2, 3, 4, 5, 6
4. Coniferous closed	5	
5. Coniferous open		3, 4
6. Mixed forest		
7. Deciduous closed	2, 4	3
8. Deciduous open		
9. Early successional	1, 2, 3	1, 2
10. Deep water	1, 2, 3, 4, 6	1, 2, 4, 5, 6
11. Shallow water	4, 6	2, 6
12. Swamp	3	1, 2
13. Marsh[a]	5, 6	5, 6
14. Fen[b]		
15. Bog[c]		

[a] No test sites found in Scene 2.
[b] No test sites found in Scenes 1, 2, 3, 4, 5.
[c] No test sites found in Scenes 1, 3, 4, 5.

Table 5. Scene numbers in which each class equalled or exceeded 75% Producer's or Consumer's accuracy in raw data classification.

Class	Producer's accuracy> 75%	Consumer's accuracy> 75%
1. Impervious	2, 4	4
2. Pervious	—	1
3. Soil	1, 4	1, 2, 3, 4, 6
4. Coniferous closed	5	—
5. Coniferous open	—	3, 4
6. Mixed forest	—	—
7. Deciduous closed	2, 4	3
8. Deciduous open	—	—
9. Early successional	1, 2, 3	1, 2
10. Deep water	1, 2, 3, 4	1, 2, 4, 5, 6
11. Shallow water	4	2
12. Swamp	3	1, 2
13. Marsh[a]	6	5, 6
14. Fen[b]		
15. Bog[c]		

[a] No test sites found in Scene 2.
[b] No test sites found in Scenes 1, 2, 3, 4, 5.
[c] No test sites found in Scenes 1, 3, 4, 5.

two scenes differed considerably. Class means differed by up to 50dn in any given spectral band. The reason for these differences in mean brightness is unknown at this time. They could be due to view angle differences, the photographic scanning process, or variations in training data selection. As a result of such differences, it was determined that all scenes (tiles from the same photograph) must be classified separately with adequate field support in each scene. It was not possible to evaluate classification of several scenes using training data from one scene as had been desired in order to determine if future operational use of such imagery could be conducted with reduced field work.

Table 6 lists aggregated classes which resulted in equal to, or greater than, 75% accuracy in the scenes evaluated. Class 1+ 3 represents a 'bare' nonvegetated surface, class 4+ 5 is coniferous forest, class 7+ 8 is deciduous forest, and the wetland class was derived from any of the four wetland classes which were available in a given scene. These aggregated classes were selected based on error matrix analysis and also to match typical USGS Level II classes which can be extracted from Landsat TM using similar spectral bands. Mixed forest was not aggregated with another class because it was generally confused with several classes. Class 2, grass/pasture,

Table 6. Scene numbers in which each aggregated class equalled or exceeded 75% Producer's or Consumer's accuracy in raw data classification. The 'wetland' class consisted of classes 12+ 13, 12+ 15, or 12– 15 depending on the classes available to be aggregated in each scene.

Class	Producer's accuracy> 75%	Consumer's accuracy> 75%
1+ 3	2, 4	1, 2, 4, 5, 6
4+ 5	1	1, 3
7+ 8	1, 2	1, 2
Wetland	—	1, 4

was usually very poor in accuracy and too difficult to merge with other classes. It was probably a poor choice in class attribute, or errors resulted in data acquisition for this class. The results in Table 6 show some improvement in mapping of the aggregated classes over the individual classes.

4.3.2. *Transformed data*

Table 7 gives summarised results from classification of the transformed image data of Scene 2. *Principal components analysis*: PC1 and PC2 are the first two principal components. They accounted for 98% of the variance while the third PC was very noisy. Such results are typical of three-band data in the visible and near-infrared (e.g. Michele Basham May *et al.* 1997, using SPOT data). The transformations were $PC1 = 0.60B1 + 0.56B2 + 0.56B3$, and $PC2 = -0.21B1 - 0.57B2 + 0.80B3$ where B_i = band number (B1 = green, B2 = red, B3 = near-infrared). In the images of PC1, rivers were very dark, forest medium brightness, and fields very bright. This component showed variance related to overall brightness and was similar in appearance to a linearly contrast stretched visible band. The PC2 image showed most fields to be dark, the river dark, and the forest very bright. It illustrated variance related to the presence or absence of green vegetation at this time of year. A two-band classification of PC1 and PC2 resulted in slightly greater overall accuracy with the Producer's accuracy being higher and the Consumer's accuracy being lower than the raw data classification. No test pixels of classes 3 and 4 were correctly assigned resulting in 0.0% accuracy. Wetlands in particular increased significantly in Producer's accuracy from 30.8% to 41.5% for swamp (12), and 7.1% to 64.1% for bog (15). The Consumer's accuracy of swamp decreased from 89.3% to 53.9% while that for bog increased from 15.5% to 40.4%. Other classes which showed increases of more than 5% in either Producer's or Consumer's accuracy and no loss of accuracy in the other measure were mixed forest (6), open deciduous (8), early successional (9), and shallow water (11). The bare soil class (3) and open coniferous class (5) had significantly decreased Producer's and Consumer's accuracy ($\geqslant$ 19% lower), while closed coniferous (4) and deep water (10) had lower Consumer accuracies but their Producer accuracies remained essentially the same.

Spectral band ratios: the overall accuracy of classification and the accuracies of most classes were lower than for the raw data classification. Only urban grass/pasture/new crops (2) and bog (15) increased in both Producer's and Consumer's accuracy (by $\geqslant$ 20%). Open coniferous (5), open deciduous (8), and shallow water (11) were better classified in one measure and about equally worse in the other. All other classes showed declines in accuracy in one or both measures. The poor results reflect the impact of high data variance on the resulting small range of real ratios (0–5) and their conversion to a range of 0–255 integers. Ideally 16- or 32-bit encoding and subsequent classification should have been conducted, but at the time the

Table 7. Results of classification of transformed data.

Bands	Average Producer's accuracy (%)	Average Consumer's accuracy (%)	Kappa
PC1, PC2	49.5	48.5	0.46
NDVI, NDVI′	36.5	36.6	0.33
3 raw+ texture	43.1	57.9	0.39

research was conducted PC hard drives did not have large enough capacity to store such data as well as the resulting thematic maps.

Texture transformation: addition of a co-occurrence texture band to the raw data in a four-band classification did not improve overall classification accuracy from that of the raw data. Within individual classes, gains or losses in either accuracy measure were all less than 10%, indicating that the texture band did not have as strong an effect on accuracy as the spectral transformations did. Both swamp (12) and bog (15) improved slightly in Consumer's accuracy (3.2% and 4.7%, respectively) without a loss in Producer's accuracy. The principal reason for these results is that the component of the high data variance caused by film scanning resulted in additional spatial variability within most land cover types. If such film scanning can be improved, cover type texture may still be useful in classification of high resolution imagery.

4.3.3. *Evaluation of context and neural network classifiers*

Table 8 gives the average Producer's and Consumer's accuracies and the kappa coefficients for the two runs of each of these classifiers. The two context classifiers were run with 7×7 and 21×21 windows, respectively. The neural network classifiers are 1, faster, and 2, slower in terms of learning and momentum rates.

The results of both of these 'advanced' classifiers were disappointing. Neither were able to classify any of the land cover types consistently. Even deep water did not have high Producer's and Consumer's accuracy. The neural network classifier did not reach the maximum allowable global error which had been specified, nor did it produce all the desired classes. The second run, with a lower learning and momentum rate did improve the accuracy significantly, but at a cost of 60 hours of CPU time for training the network and another 8 h of classification.

4.4. *Combination of processing methods*

As a result of the various processes applied to the Scene 2 data, it became evident that large improvements in classification accuracy were not going to be achieved easily. PCA provided the best overall improvement, but several classes were very poorly classified. Texture processing gave slightly poorer results but some classes were improved in either Producer's or Consumer's accuracy. In addition, it was decided to introduce noise reduction filtering at this point, although such a process was avoided in initial analysis because it defeated the goal of evaluation of the high resolution capabilities of the MK-4 imagery. A test of a 5×5 median filter revealed that soil (3) was improved by 5–7% in both accuracy measures while closed coniferous (4) and open coniferous (5) improved by 5.4% and 12.0% in Producer's accuracy

Table 8. Results of classification of context and neural network classifiers applied to raw data of Scene 2.

Classifier	Average Producer's accuracy (%)	Average Consumer's accuracy (%)	Kappa
Context (7×7)	29.1	34.1	0.24
Context (21×21)	28.3	24.0	0.22
Neural Network$_1$	12.6	9.3	0.12
Neural Network$_2$	21.0	11.5	0.20

without a loss in Consumer's accuracy. Two classes, closed deciduous (7) and bog (15) had significantly decreased accuracy in both measures.

By examining the processes of PCA, addition of texture and noise reduction, it became evident that each process improved the accuracy of a different set of classes. PCA improved classes 6, 8, 9, 11 and 15. Addition of texture improved (slightly) classes 12 and 15. Noise reduction improved classes 2, 3, 4, 5 and 8. The only classes not improved by one of these processes were 1, 7 and 10. Of these, each had decreased accuracy in only one of the Producer's or Consumer's accuracy measures in only one of the processes of PCA, addition of texture or noise reduction. Thus, it was felt that a combination of these processes may improve overall classification accuracy by accumulating improvement in individual class accuracies from any one or more of the processes. It was therefore decided to first process the three raw spectral bands using the 5×5 median filter. These images were used in PCA and the first two components were extracted. The texture program was run on PC2. PC2 was used because it visually appeared to contain much more useful texture information than any of the original bands. It was also hypothesized that texture of an image which had been produced from noise reduced data would not be as affected by random noise.

The final input to classification was therefore three channels: PC1, PC2 and texture of PC2, all being derived from initially noise reduced data. The results were a kappa of 0.42, average Producer's accuracy of 45.5%, and average Consumer's accuracy of 52.7%. Twelve classes decreased in accuracy while 10 classes increased in accuracy. Such lack of distinct improvement after a significant amount of data processing demonstrated that there was not a straightforward method for improving the map accuracy over that derived from raw data.

5. Discussion

This research has evaluated the thematic mapping potential of a recently available type of high-resolution multispectral satellite imagery using a variety of data processing and classification techniques. Although the point-based accuracy of all the maps was low for most classes, the maps visually appeared to be similar to those typically produced using Landsat or SPOT. Many fairly uniform areas such as agricultural fields had a majority of pixels classified correctly so the rectangular geometry of the fields was quite evident. The accuracy of some classes was in the range of those found for classes of similar detail using SPOT and conventional classification techniques (e.g. Kontoes and Rokos 1996, Laba *et al.* 1997, Michele Basham May *et al.* 1997) or even of more advanced sensors and classification techniques (e.g. Ustin *et al.* 1996 using AVIRIS and spectral mixture modelling). Finally, as a result of the small pixel spacing of the imagery, more small features were distinguished than is possible with lower resolution imagery, resulting in maps with quite large variations in polygon sizes. For example, Parks Canada has found the maps to be useful in identification of small, previously noninventoried wetlands (the most critical class in environmental management of the region) which require further field examination. However, despite these positive results, it was still obvious that most polygons contained a significant proportion of individual pixels which were misclassified and which contributed to the low point-based accuracy. Such thematic maps should not be relied upon for pixel-based measurement and Geographical Information System (GIS) operations such as class inventory or overlay operations without modelling and correction of resulting errors (e.g. using regression of known patch size *versus* mapped patch size, Hlavka and Livingstone 1997).

In carrying out the research, four key issues were identified which will be applicable to thematic mapping using future high resolution multispectral satellite data. They are related to: (1) image data acquisition; (2) field data acquisition; (3) a need for alternatives to pixel-based mapping and accuracy assessment techniques; and (4) a need for land cover class attribute schemes which are appropriate for such high spatial variance data.

In image data acquisition, an obvious cause of subsequent low mapping accuracy was the high data variance which appeared to consist of both noise from the image acquisition process and of high spatial variance in land cover spectral characteristics. These two variance components were difficult to separate when visually viewing the imagery. In future high resolution satellite imaging with solid-state sensors, the signal noise is not expected to be as high as in this scanned photographic data. In addition, film scanning quality continues to improve. Both the random and systematic noise evident in this imagery would be significantly reduced using current scanning technology. A second obvious cause of low mapping accuracy was the difference in time between the image acquisition date and the field work (Congalton and Green 1993, Hammond and Verbyla 1996). This is very common with optical satellite imagery—acquisition of desired imagery at an optimal time for mapping of targets of interest has always been an ideal that is not often achieved. Thus, the data of this study were considered to give more of an indication of the potential utility of such imagery under operational conditions and attempts were not made to acquire more suitable data until after the study had been completed.

In field data acquisition and verification, two factors affecting error emerged. First, clusters of training and test pixels were sampled at each site. While random sampling of individual pixels is statistically more powerful (Miller *et al.* 1988, Gong and Howarth 1990, Dobbertin and Biging 1996), it is virtually impossible to conduct practically using such high-resolution imagery as identification of exact pixel locations on the ground and their land cover constituents is very difficult. In addition, in this study, training and test samples were taken which were representative of both the large size distribution and spectral variation of each class since there were large ranges of spatial extents and surface cover influences on spectral reflectance of most classes. This results in a conservative estimate of accuracy in contrast to the common technique of sampling within only large homogeneous areas of minimum spectral variance which can skew the accuracy estimates upwards (Hammond and Verbyla 1996). Second, in conventional mapping with Landsat and SPOT at USGS Levels 1 and 2, the land cover types are broad and have large spatial extents. The minimum mapping unit is typically on the order of 2×2 pixels or larger representing a minimum area on the ground of $3600\,m^2$, or 0.36 ha for Landsat (e.g. as in Michele Basham May *et al.* 1997, Miller *et al.* 1998). With such a minimum mapping unit of 2×2 pixels in MK-4 data, the equivalent area on the ground is only $225\,m^2$, or 0.023 ha. Navigating to pre-selected sites or matching GPS co-ordinates to image co-ordinates requires much more precision than for coarser resolution data. The best relative error between image and ground co-ordinates achieved in this study was on the order of 10 m. Verbyla and Hammond (1995) associate such increased relative positional error with decreased classification accuracy. As a consequence of these characteristics, field data acquisition and verification for mapping with future satellites of 1–5 m pixel spacing can be expected to increase in difficulty and cost.

Point-based classification techniques do not appear to be suitable for regional thematic mapping using such high variance data. Neither the original raw data, nor data transformations produced acceptable results for all classes. This agrees with other studies of simple spectral transformations in land cover mapping. Anji Reddy and Reddy (1996) reported that two channel classification of the blue/green and red/near-infrared ratios of IRS data produced significantly less accuracy than the four raw spectral bands. Michele Basham May *et al.* (1997) found no significant change in accuracy of classes of similar detail to those of this study using PCA of TM and SPOT data. In contrast to this study though, most studies of advanced point-based classifiers such as neural networks have found significant improvement in classification accuracy (Peddle *et al.* 1994, Bruzzone *et al.* 1997, Yool 1998). The methods used in this research were the simplest available as the objective was to minimize user input to determine potential for transfer to nonexperts. Perhaps the neural network performance would have improved given more interactive modification of learning and momentum rates (Yool 1998), greater numbers of independent training pixels or greater numbers of discriminating variables (Foody and Arona 1997), modification of the internal network structure (Paola and Schowengerdt 1997), use of 'soft' mixture classes (Foody 1996, Bernard *et al.* 1997) or a variety of other variations of neural network techniques and input parameters (Kanellopoulos and Wilkinson 1997). However, Skidmore *et al.* (1997) in a sobering evaluation of the vast range of possible inputs and structures concluded that the impact of each is quite significant and too unpredictable for any given dataset and that the use of neural networks in operational land cover mapping will be limited. As an alternative to conventional point-based classification, research with SPOT and Landsat data has shown that incorporation of image texture (Peddle and Franklin 1991, Gong *et al.* 1992, Dreyer 1993, Augusteijn *et al.* 1995, Dikshit 1996, Bruzzone *et al.* 1997) can improve classification accuracy although most results, as in this study, showed only marginal improvements. Other classification techniques which have proven successful with Landsat or SPOT and which have good potential for future high resolution satellite mapping include: (1) use of image context such as spatial autocorrelation (e.g. Brown and Walsh 1993, Flygare 1997), geographic context such as environmental information (e.g. Kontoes and Rokos 1996), or shape information (Li 1996); (2) object-based classification (e.g. Janssen *et al.* 1990, Fung and Chan 1994, Lobo *et al.* 1996) where the objects of interest such as fields, forest stands, or water bodies are first segmented in some way and then evaluated as whole entities; (3) knowledge-based classification (e.g. Peddle 1995) where expert rules or judgements may be made which aid the class discrimination process; and (4) visual classification as Martin and Howarth (1989) found that interpreter accuracy of SPOT data was significantly better than classification using automated techniques in the complex environment of the rural–urban transition. Each of these will need thorough testing as the small subset of context and texture analysis conducted in this research did not produce positive results. As with the classification process, in assessment of accuracy, the vast proportion of studies have used point-based techniques. However, when there is such high variability from pixel to pixel, a more appropriate approach would be object-based accuracy assessment (e.g. using majority, most frequent or fuzzy rules to determine if an object is mostly correct). Parks Canada conducts such an assessment informally when studying the thematic maps produced by this research to determine if an area is actually the class which is on the map. If a majority of pixels indicate a given class of interest, subsequent field work verifies whether the object class is correct.

In class attribute definition, the principal issue of the potential applicability of small pixel size imagery in regional thematic mapping must be addressed. The results obtained in this study provide evidence that land cover classification will be hindered by high spatial data variance if attributes are used which are too coarse in detail. Land cover attributes which have spatial extents representative of the image spatial variance must be defined. In this research, variations at the 7.5 m pixel scale in the amount and type of vegetative cover, and in soil moisture affected the classification of agricultural fields, early successional or transitional vegetation classes, and urban or managed grass classes. Variations in gap size and distribution, and in species composition at the tree level affected classification of the forest and wetland classes. Variations in urban objects such as cars, roofs, and sidewalks affected the urban class. Class attributes at USGS Levels 3 and 4 were probably more appropriate for this application, and use of 'soft' classification of mixture classes would allow pixels to be classified within the continuum of spectral variation between classes. However, the number of possible attributes at these detailed classification levels is very high so it was impossible to select a suitable subset for this unknown image data type. It was also too difficult to determine *a priori* which detailed classes could be effectively used in environmental management since such detail had never been incorporated into decision making (nor was it necessarily desired at this point in time) at regional scales. As a consequence, the high spatial variability of this data, and of anticipated higher resolution satellite data, may hinder regional land mapping and management if conventional techniques are applied. More research is needed to match management information requirements with image data sources and mapping techniques.

6. Conclusions

Russian MK-4 scanned multispectral photography has been evaluated in thematic mapping of land cover classes in Eastern Ontario using a variety of data processing and classification techniques which incorporate spectral and spatial information. The point-based accuracy of the thematic maps was generally low for most classes. Improvements in the data imaging date, data processing and classification design may increase class accuracy to levels which are useful in operational mapping. Many of the operational issues encountered in this study may be indicative of what can be expected with future high resolution satellite imaging.

Acknowledgements

Parks Canada is gratefully acknowledged for funding this research. Mark van Kralingen and Reed Kirkpatrick provided critical field assistance.

References

Anderson, J. R., Hardy, E. E., Roach, J. T., and Witmer, R. E., 1976, A land use and land cover classification system for use with remote sensor data. US Geological Survey, Professional Paper 964, Washington, DC, 28.

Anji Reddy, M., and Reddy, K. M., 1996, Performance analysis of IRS-bands for land use/land cover classification system using Maximum Likelihood Classifier. *International Journal of Remote Sensing*, **17**, 2505–2515.

Augusteijn, M. F., Clemens, L. E., and Shaw, K. A., 1995, Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 616–625.

Bernard, A. C., Wilkinson, G. G., and Kanellopoulos, I., 1997, Training strategies for neural network soft classification of remotely-sensed imagery. *International Journal of Remote Sensing*, **18**, 1851–1856.

Brown, D. G., and Walsh, S. J., 1993, Spatial autocorrelation in remotely-sensed and GIS data. *Proceedings ASPRS-ASCM Annual Convention, New Orleans LA, 15–18 February, 1993* (Bethesda, MD: American Society for Photogrammetry and Remote Sensing), pp. 31–39.

Bruzzone, L., Conese, C., Maselli, F., and Roli, F., 1997, Multisource classification of complex rural areas by statistical and neural-network approaches. *Photogrammetric Engineering and Remote Sensing*, **63**, 523–533.

Civco, D. L., 1993, Artificial neural networks for land-cover classification and mapping. *International Journal of Geographical Information Systems*, **7**, 173–186.

Congalton, R. G., 1991, A review of the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, **37**, 35–46.

Congalton, R. G., and Green, K., 1993, A practical look at the sources of confusion in error matrix generation. *Photogrammetric Engineering and Remote Sensing*, **59**, 641–644.

Dikshit, O., 1996, Textural classification for ecological research using ATM images. *International Journal of Remote Sensing*, 17, 887–915.

Dobbertin, M., and Biging, G. S., 1996, A simulation study of the effect of scene autocorrelation, training sample size and sampling method on classification accuracy. *Canadian Journal of Remote Sensing*, **22**, 360–367.

Dreyer, P., 1993, Classification of land cover using optimised neural nets on SPOT data. *Photogrammetric Engineering and Remote Sensing*, **59**, 617–621.

Flygare, A.-M., 1997, A comparison of contextual classification methods using Landsat TM. *International Journal of Remote Sensing*, **18**, 3835–3842.

Foody, G. M., 1996, Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. *International Journal of Remote Sensing*, **17**, 1317–1340.

Foody, G. M., and Arora, M. K., 1997, An evaluation of some factors affecting the accuracy of classification by an artificial neural network. *International Journal of Remote Sensing*, **18**, 799–810.

Fung, T., and Chan, K.-C., 1994, Spatial composition of spectral classes: a structural approach for image analysis of heterogeneous land-use and land-cover types. *Photogrammetric Engineering and Remote Sensing*, **60**, 173–180.

Gong, P., and Howarth, P. J., 1990, An assessment of some factors influencing multispectral land-cover classification. *Photogrammetric Engineering and Remote Sensing*, **56**, 597–603.

Gong, P., and Howarth, P. J., 1992, Frequency-based contextual classification and grey-level vector reduction for land-use identification. *Photogrammetric Engineering and Remote Sensing*, **58**, 423–437.

Gong, P., Marceau, D. J., and Howarth, P. J., 1992, A comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data. *Remote Sensing of Environment*, **40**, 137–151.

Hammond, T. O., and Verbyla, D. L., 1996, Optimistic bias in classification accuracy assessment. *International Journal of Remote Sensing*, **17**, 1261–1266.

Hansen, A. R., 1994, Change detection analysis using Landsat Thematic Mapper data and Landsat Multispectral Scanner data: training site definition for the St Lawrence Islands National Park study area. Report to St Lawrence Islands National Park, Parks Canada, 48pp.

Hlavka, C. A., and Livingstone, G. P., 1997, Statistical models of fragmented land cover and the effect of coarse spatial resolution on the estimation of area with satellite sensor imagery. *International Journal of Remote Sensing*, **18**, 2253–2259.

Janssen, L. L. F., Jaarsma, M. N., and van der Linden, E. T. M., 1990, Integrating topographic data with remote sensing for land-cover classification. *Photogrammetric Engineering and Remote Sensing*, **56**, 1503–1506.

Jensen, J. R., 1996, *Introductory Digital Image Processing*, 2nd edn (Toronto: Prentice-Hall).

Kanellopoulos, I., and Wilkinson, G. G., 1997, Strategies and best practice for neural network image classification. *International Journal of Remote Sensing*, **18**, 711–725.

Kontoes, C. C., and Rokos, D., 1996, The integration of spatial context information in an experimental knowledge-based system and the supervised relaxation algorithm—two successful approaches to improving SPOT-XS classification. *International Journal of Remote Sensing*, **17**, 3093–3106.

Laba, M., Smith, S. D., and Degloria, S. D., 1997, Landsat-based land cover mapping in the lower Yuna River watershed in the Dominican Republic. *International Journal of Remote Sensing*, **18**, 3011–3025.

Li, X., 1996, A method to improve classification with shape information. *International Journal of Remote Sensing*, **17**, 1473–1481.

Lobo, A., Chic, O., and Casterad, A., 1996, Classification of Mediterranean crops with multi-sensor data: per-pixel *versus* per-object statistics and image segmentation. *International Journal of Remote Sensing*, **17**, 2385–2400.

Maren, A. J., Harston, C. T., and Pap, P. M., 1980, *Handbook of Neural Network Computing Applications* (San Diego, CA: Academic Press).

Martin, L. R. G., and Howarth, P. J., 1989, Change detection accuracy assessment using SPOT multispectral imagery of the rural-urban fringe. *Remote Sensing of Environment*, **30**, 55–66.

Martin, L. R. G., Howarth, P. J., and Holder, G. H., 1988, Multispectral classification of land-use at the rural-urban fringe using SPOT data. *Canadian Journal of Remote Sensing*, **14**, 72–79.

Michele Basham May, A., Pinder III, J. E., and Kroh, G. C., 1997, A comparison of Landsat Thematic Mapper and SPOT multi-spectral imagery for the classification of shrub and meadow vegetation northern California, U.S.A. *International Journal of Remote Sensing*, **18**, 3719–3728.

Miller, A. B., Bryant, E. S., and Birnie, R. W., 1998, An analysis of land cover changes in the Northern Forest of New England using multi-temporal Landsat MSS data. *International Journal of Remote Sensing*, **19**, 245–265.

Moody, A., and Woodcock, C. E., 1994, Scale-dependent errors in estimation of land cover proportions: Implications for global land cover datasets. *Photogrammetric Engineering and Remote Sensing*, **60**, 585–594.

Narendra, P. M., and Goldberg, M., 1977, A non-parametric clustering scheme for Landsat. *Pattern Recognition*, **9**, 207–215.

Paola, J. D., and Schowengerdt, R. A., 1997, The effect of neural-network structure on a multispectral land-use/land-cover classification. *Photogrammetric Engineering and Remote Sensing*, **63**, 535–544.

Peddle, D. R., 1995, Knowledge formulation for supervised evidential classification. *Photogrammetric Engineering and Remote Sensing*, **61**, 409–417.

Peddle, D. R., and Franklin, S. E., 1991, Image texture processing and data integration for surface pattern discrimination. *Photogrammetric Engineering and Remote Sensing*, **57**, 413–420.

Peddle, D. R., Foody, G. M., Zhang, A., Franklin, S. E., and LeDrew, E. F., 1994, Multisource image classification II: an empirical comparison of evidential reasoning and neural network approaches. *Canadian Journal of Remote Sensing*, **20**, 396–407.

Skidmore, A. K., Turner, B. J., Brinkof, W., and Knowles, E., 1997, Performance of a neural network: mapping forests using GIS and remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, **63**, 501–514.

Tou, J. T., and Gonzalez, R. C., 1977, *Pattern Recognition Principles* (Reading, MA: Addison-Wesley).

Treitz, P. M., Howarth, P. J., and Gong, P., 1992, Application of satellite and GIS technologies in land cover and land use mapping at the rural–urban fringe: a case study. *Photogrammetric Engineering and Remote Sensing*, **58**, 439–448.

Ustin, S. L., Hart, Q. J., Duan, L., and Scheer, G., 1996, Vegetation mapping on hardwood rangelands in California. *International Journal of Remote Sensing*, **17**, 3105–3036.

Verbyla, D. L., and Hammond, T. O., 1995, Conservative bias in classification accuracy assessment due to pixel-by-pixel comparison of classified images with reference grids. *International Journal of Remote Sensing*, **16**, 581–587.

Yool, S. R., 1998, Land cover classification in rugged areas using simulated moderate-resolution remote sensor data and an artificial neural network. *International Journal of Remote Sensing*, **19**, 85–96.

Yuan, X., King, D. J., and Vlcek, J., 1991, Sugar maple decline assessment based on spectral and textural analysis of multispectral aerial videography. *Remote Sensing of Environment*, **37**, 47–54.