INT. J. REMOTE SENSING, 2004, VOL. 25, NO. 12, 2421–2440

Contextual classification of Landsat TM images to forest inventory cover types

S. MAGNUSSEN, P. BOUDEWYN and M. WULDER

Natural Resources Canada, Pacific Forestry Center, 506 West Burnside Road, Victoria, British Columbia, Canada V8Z 1M5; e-mail: Steen.Magnussen@nrcan.gc.ca

(*Received 7 January 2002; in final form 21 August 2002*)

Abstract. In the context of Landsat TM images forest stands are a cluster of homogeneous pixels. Contextual classification of forest cover types exploits relationships between neighbouring pixels in the pursuit of an increase in classification accuracy. Results with six contextual classifiers from two sites in Canada were compared to results with a maximum likelihood (ML) classifier. The comparisons were done at three levels of spectral class separation. Training and validation data were obtained from single-stage cluster sampling of 2 km × 2 km primary sampling units (PSU) located on a 20 km × 20 km grid. A strong relationship between contextual and ML classification accuracy was explored with logistic regression analysis. Effects of contextual classification were predicted for given levels of ML accuracy. Estimates of the spatial autocorrelation of reflectance values within a PSU were deemed consistent with a first-order autoregressive process. Iterative Conditional Modes (ICM) was the best contextual method; it improved the overall accuracy by four to six percentage points (statistically significant) when ML accuracy was between 50% and 80%. A relaxed ICM and a smoothing algorithm were second and third best. Contextual classification is most promising when an ML accuracy is around 70%. ICM results were sensitive to the level of spatial autocorrelation of ML classification errors and to the homogeneity of a PSU.

1. Introduction

Forest cover type maps are composed of polygons or forest stands with a content, structure and composition matching a type description. In a Landsat TM image with a nominal pixel size of 30 m a forest stand occupies anywhere from a few to several hundred or even thousands of pixels. Spectral reflectance values of pixels in close spatial proximity within a given stand tend to be similar. Conversely, the prevalence of spatially contiguous stands improves the likelihood that adjoining image pixels belong to the same cover type class. Contextual classification exploits these relationships among neighbouring pixels as opposed to a per-pixel classification which derives a cover type from the information linked to a single pixel. Stable, consistent, and predictable relationships among neighbouring pixels can be quantified and used to improve the classification accuracy. Switzer (1980) pioneered the use of spatial smoothing in a classification context. Checkered agricultural landscapes were especially amenable to contextual classification (Goldberg and Goodenough 1978). Since then a large array of contextual methods

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160310001642296

derived from Markov random fields, spatial statistics, Bayesian methods, fuzzy logic, segmentation, texture, or neural nets have emerged. When tried on fixed binary images degraded by random noise, these methods have been very effective at restoring the original image (Ripley 1985). Practical experience with these methods under realistic conditions in a forest inventory context remains scant, but published studies indicate an increase in accuracy after including context in the classification schemes (Palubinskas *et al.* 1995, Slaymaker *et al.* 1996, Flygare 1997, Stuckens *et al.* 2000, Hubert-Moy *et al.* 2001).

Context in this study is restricted to mean a neighbourhood of pixels. Integrating information from neighbouring pixels into a classification scheme is more complex and time-consuming than a per-pixel classification. These efforts have to be justified by a commensurate increase in accuracy. To assess whether contextual classification is an attractive option for forest type classification of satellite images, this study compared the performance of six contextual methods to a maximum likelihood per-pixel classification in two contrasting forested landscapes, and at three levels of spectral class separation. Our approach to statistical inference is design-based (Gregoire 1998). Results are obtained from a random sample of 2 km × 2 km primary sample units.

2. Material and methods

2.1. *Study sites*

Contextual and per pixel classification accuracy was compared with data from two study sites, one in New Brunswick and one in Saskatchewan (figure 1). The forested landscape in the New Brunswick site appears more fragmented than the Saskatchewan site due to a larger population density and longer history of anthropogenic influence. Forest stand polygons in the New Brunswick site have a mean size of 5.2 ha compared to 6.5 ha in Saskatchewan. Both study sites are relatively flat (average slope was 2% in New Brunswick and 0.6% in Saskatchewan) with an inter-quartile altitude range of 70–120 m in New Brunswick and 500–600 m in Saskatchewan.

2.2. *Landsat TM image data*

Landsat TM images used for classification are detailed in table 1. The TM scenes were co-registered to distinct ground control points (37 in New Brunswick and 86 in Saskatchewan). Ground control points were transferred from a 1 : 50 000 National Topographic Series map (New Brunswick) and 1 : 12 500 ortho-rectified aerial photos (Saskatchewan) to their respective Landsat scene. Estimated rms positional error determined with a second-order polynomial and cubic resampling was 0.46 and 0.76, respectively (unit: pixel). Channels used were 2, 3, 4, 5, and 7. Atmospheric corrections were done according to Richter (1990). Correlation of reflectance values between TM channels was reduced to zero via a transform based on a Cholesky decomposition of the covariance matrix of the reflectance data (Cressie 1991: 202–203). Visual inspection of histograms of reflectance data indicated that the assumption of a multivariate Gaussian distribution was tenable. Yet, formal testing (D'Agostino *et al.* 1990) rejected the null hypothesis of normality of the transformed reflectance data due to an excess of kurtosis in all channels ($p < 0.01$).

Figure 1. Location map of study sites (black boxes) in New Brunswick and Saskatchewan.

Table 1. Landsat TM data.

Landsat TM scene	New Brunswick	Saskatchewan
Acquisition date	July 1995	July 1994
Track	9	37
Frame	28	22

2.3. *Forest inventory data*

Forest cover type maps obtained from interpretation of aerial photography served as ground truth for the training and validation of the classifiers. Years of photography were 1993–1994 in New Brunswick (New Brunswick Department of Natural Resources and Energy), and 1994–1996 in Saskatchewan (Forest Management Branch and Saskatchewan Environment and Resource Management, Forestry Branch, Inventory Unit). Cover types and their definitions are listed in table 2.

Visual inspection of an overlay of TM and inventory data identified clusters of pixels where the inventory cover type label was dated due to a significant change since the inventory was completed. We masked these pixels ($\sim 1\%$) from the analyses. TM reflectance values deviating more than three standard deviations from their cover type conditional mean ($< 1\%$) were also masked as they were visible outliers.

Table 2. Forest inventory cover types.

Cover type	Definition
Exposed land (el)	Non-vegetated polygons (trees, shrubs, herbs, or bryoids cover less than 5% of the polygon). This type includes features such as roads, buildings, exposed soil, river sediments, and rock outcrops.
Shrubs (sh)	Vegetation covers at least 5% of the polygon, and is comprised mainly of shrub species of any height (woody perennial plants that have a relatively low growth habit and non-treelike form).
Treed, broad-leaved (tb)	Trees cover at least 10% of the polygon, and are comprised mainly of deciduous species (75% or more of total tree basal area).
Treed, coniferous (tc)	Trees cover at least 10% of the polygon, and are comprised mainly of coniferous species (75% or more of total tree basal area).
Treed, mixed-wood (tm)	Trees cover at least 10% of the polygon, and are comprised of both deciduous and coniferous species (less than 75% of total tree basal area in each species group).
Water (wa)	Lakes, rivers, streams and reservoirs.
Crop/Pasture (cp)	Cultivated land, fields, pastures, and agricultural land.

2.4. *Training and validation data*

Training (fitting) and validation of classifiers was done within the context of a one-stage cluster sampling (Cochran 1977) with 2 km × 2 km primary sampling units (*PSU*) on a 20 × 20 km grid (Magnussen *et al.* 1998). Each PSU contained approximately 4300 30 m × 30 m co-registered pixels (secondary sampling units) from the TM image and the rasterized forest cover type maps. The New Brunswick site had 57 PSUs and the Saskatchewan 81 PSUs. The PSUs on each site were split equally into one of three sets (SET1, SET2, SET3) by random assignments. Each set was used once for training with the remaining two site-specific sets serving validation. Thus the scheme generated three sets of validation results.

2.5. *Three levels of spectral class separation*

The effect of contextual classification was expected to depend on the spectral class separation. With a low separation the confusion of classes will be high and a contextual classification may only degrade an already low accuracy, and vice versa for a high separation. The results of a comparison of contextual and per-pixel classification would therefore depend on the spectral separation of the adopted cover types. To generalize our results we manipulated the TM reflectance data to obtain three levels of class separation (SEP1, SEP2, and SEP3). SEP1 corresponds to the original TM data. Datasets SEP2 and SEP3 were generated by adding a class-dependent offset to TM channels 2 and 3. Offsets for SEP2 and SEP3 were chosen to give a minimum Mahalanobis distance (McLachlan 1991) in class mean reflectance of 1.25 and 1.5, respectively. Offsets are listed in table 3 and the corresponding Mahalanobis class mean distances are given in table 4.

2.6. *Maximum likelihood classification (ML)*

All pixels in a validation set were classified to a cover type class (C, $C = 1, \ldots, K$) by an ML classifier. With the assumption of multivariate normal (transformed) reflectance data, the class predicted for pixel i was the class with the highest

Table 3. Offsets added to TM channel 2 and 3 reflectance values to achieve a minimum Mahalanobis distance between class mean reflectance values of 1.25 (SEP2), and 1.50 (SEP3).

		SEP2		SEP3	
		Channel 2	Channel 3	Channel 2	Channel 3
New Brunswick	Exposed land	0.75	3.53	1.29	6.09
	Shrubs	−0.06	1.36	−0.11	−2.31
	Treed, broad-leaved	−1.36	−1.25	−2.37	−2.16
	Treed, coniferous	−0.15	0.94	−0.26	1.63
	Treed, mixed-wood	−0.15	−0.35	−0.26	−0.61
	Water	1.37	−3.02	2.37	−5.21
	Crop/pasture	−0.38	1.50	−0.66	2.58
Saskatchewan	Exposed land	−0.14	3.15	−0.26	5.90
	Shrubs	0.01	−0.85	0.01	−1.59
	Treed, broad-leaved	−0.64	−0.87	−1.19	−1.63
	Treed, coniferous	−0.46	−1.60	−0.86	−2.99
	Treed, mixed-wood	−0.46	−1.60	−0.86	−2.99
	Water	1.55	0.95	2.91	1.77

Table 4. Mahalanobis distances of mean spectral reflectance between two cover type classes at three levels of class separation (SEP1, SEP2, SEP3). Cover type abbreviations are defined in table 2.

	New Brunswick			Saskatchewan		
Class pair	SEP1	SEP2	SEP3	SEP1	SEP2	SEP3
el, sh	1.73	5.42	8.49	2.28	4.50	7.06
el, tb	2.46	6.32	9.72	2.93	4.78	7.34
el, tc	2.38	3.87	5.44	2.89	4.50	6.87
el, tm	2.26	4.81	7.22	2.48	5.23	8.39
el, wa	5.81	8.83	12.15	6.47	6.76	7.51
el, cp	1.35	3.01	4.53	n.a.	n.a.	n.a.
sh, tb	0.91	1.81	2.68	1.93	2.05	2.30
sh, tc	1.29	2.56	3.83	1.69	1.72	1.80
sh, tm	0.81	1.25	1.75	0.89	1.24	1.83
sh, wa	5.31	5.83	6.51	6.14	7.28	8.66
sh, cp	1.49	3.11	4.70	n.a.	n.a.	n.a.
tb, tc	1.30	3.08	4.72	2.31	2.36	2.44
tb, tm	0.64	1.93	3.02	1.39	1.63	1.95
tb, wa	5.59	6.67	8.03	6.73	8.02	9.71
tb, cp	2.17	3.82	5.55	n.a.	n.a.	n.a.
tc, tm	0.68	1.44	2.18	1.23	1.65	2.14
tc, wa	4.46	6.16	8.11	4.63	6.05	7.77
tc, cp	2.66	2.71	2.81	n.a.	n.a.	n.a.
tm, wa	5.03	5.95	7.11	5.84	7.61	9.73
tm, cp	2.25	2.87	3.67	n.a.	n.a.	n.a.
wa, cp	6.29	7.91	9.92	n.a.	n.a.	n.a.
mean	2.71	4.26	5.82	3.32	4.36	5.70

conditional log likelihood (LL). Specifically, the cover type class C_i predicted for pixel i in a validation set was the class that minimized

$$-LL_i(C_i=k)=\frac{1}{2}(\mathbf{y}_i-\tilde{\boldsymbol{\mu}}_k)'\tilde{\boldsymbol{\Sigma}}_k^{-}(\mathbf{y}_i-\tilde{\boldsymbol{\mu}}_k)+\frac{1}{2}\log|\tilde{\boldsymbol{\Sigma}}_k|,\ k=1,\ldots,K, \tag{1}$$

where $\mathbf{y}_i$ is the vector of transformed reflectance values for pixel i, $\tilde{\boldsymbol{\mu}}_k$ is the generalized mean of transformed reflectance values for class k, and $\tilde{\boldsymbol{\Sigma}}_k$ is the generalized covariance matrix of transformed reflectance values for class k. A′ denotes a transpose of a vector or a matrix, $\mathbf{W}^-$ is the generalized inverse of $\mathbf{W}$, and $|\mathbf{W}|$ is the determinant of $\mathbf{W}$. Class dependent means and covariance were obtained from the training data. The cluster sampling of training data necessitates the use of generalized means, variances and covariance, as ordinary maximum likelihood estimates of the same would ignore the spatial covariance of observations in a PSU (Cressie 1991), resulting in a biased classification (McLachlan 1991). Details on estimation of the generalized means, variance and covariance are in the Appendix. The multivariate normal density function implicit in (1) was chosen despite significant departures from normality in the training data. However, since a non-parametric classifier (McLachlan 1991) gave no improvement we opted for the multivariate normal model due to simplicity of computation.

2.7. *Contextual classifiers*

2.7.1. *Iterated Conditional Modes (ICM)*

Besag (Besag 1986) formulated a pseudo likelihood of images based on a locally dependent Markov random field (MRF) in which, *a priori*, the class membership probability of a pixel i is proportional to the number of pixels with the same class prediction in a nearest neighbour clique (NN_i) surrounding pixel i. In ICM all pixels, in a randomly chosen sequence, are reclassified to the class that minimizes the energy function E (Geman and Geman 1984) in (2). This process is iterated until the number of pixels that change class between two iterations has become negligible:

$$E\left(C_i=k\middle|\mathbf{C}^j_{NN_i}\right)=-LL_i(C=k)-\beta_k\times\eta\left(k,C^j_{NN_i}\right),\ k=1,\ldots,K,\ j=0,1,2,3,\ldots, \quad (2)$$

where $\mathbf{C}^j_{NN_i}$ is the set of class predictions for NN_i after the jth iteration of ICM, $\eta\left(k,\mathbf{C}^j_{NN_i}\right)$ is the number of class k predictions in $\mathbf{C}^j_{NN_i}$, and β_k is a class-specific 'temperature constant' estimated from the training data. ICM was stopped when less than 0.02% of the pixels changed class during an iteration. The median number of iterations was 3 (min 2, max 6). $\mathbf{C}^0_{NN_i}$ was obtained from the initial ML classifications. NN_i pixels located outside a PSU were imputed by a wrap-around of the PSU on a torus (Ripley 1985).

According to the ICM model we have (McLachlan 1991: 430 (equation 13.6)):

$$\text{prob}\left(C_i=k|\mathbf{C}^j_{NN_i}\right)\propto\text{Exp}\left[\beta_k\times\eta\left(k,C^j_{NN_i}\right)\right],\ k=1,\ldots,K,\ j=0,1,2,3,\ldots \quad (3)$$

Estimates of β_k were obtained by ordinary least squares linear regression with the logarithm of the relative frequency of class k pixels in the training data that have n_k class k nearest neighbours as the dependent variable and n_k as the predictor. Relative frequencies of zero were replaced by a value of $10^{-2.61}$ which is -6.0 on a logarithmic scale. Estimates of β_k varied from 0.28 to 0.61 with a mean of 0.48.

2.7.2. *Modified Iterative Conditional Modes (MICM)*

Chou and Brown (1990) and others criticized ICM for heavy dependence on the initial classification. They proposed modifications towards a more flexible classification that takes into consideration the degree of risk involved in classification. Van Deusen (1995) proposed a modified ICM which he called 'modified

highest confidence first classification' that we shall call MICM. MICM introduces an 'energy differential' (ΔE) to quantify the uncertainty of a classification. The energy differential of a pixel i is defined as the difference in energy (e.g. equation (2)) between the class with the lowest energy and the class with the next lowest energy. Pixels with a low ΔE have uncertain class predictions, and pixels with an ΔE below a certain threshold, say ΔE_c, should not be allowed to influence the class prediction of a neighbouring pixel. The MICM classification followed the steps recommended by Van Deusen:

1. Calculate ΔE from equation (2) with $\beta_k = 0$ for all pixels. Equate ΔE_c for a given validation set to the 30th percentile of the ΔE values in the corresponding training dataset.
2. ML-classify all pixels with $\Delta E \geqslant \Delta E_c$, remaining pixels are classified as 'unknown'.
3. ICM-reclassify pixels with an energy differential computed for the current iteration (ΔE^i) larger or equal to ΔE_c as outlined in §2.4.1 but discard pixels labelled as 'unknown' in step 2 from any nearest neighbour clique.
4. Increase β_k to 1.0 for all classes and repeat step 3.
5. With β_k fixed at 1.0 lower ΔE_c to $\Delta E_c \times 2^{-1}$ and repeat step 3.
6. With β_k fixed at 1.0 lower ΔE_c to $\Delta E_c \times 4^{-1}$ and repeat step 3.
7. *ML*-classify all remaining pixels with class 'unknown'

Step 3 excludes pixels with uncertain class predictions from influencing the classification of a neighbour, and it can raise the energy differential of some pixels based on nearest neighbour information to levels above the threshold which would trigger an ICM cover type prediction. Steps 4–7 are designed to gradually reclassify all pixels in the unknown category in order of prediction certainty.

2.7.3. *Directional Neighbourhood Approach (DNA)*

Press (1996) proposed a dynamic data-driven choice of neighbour pixels to be included as predictors for pixel i (i.e. NN_i). A fixed NN_i for all pixels can be sub-optimal in boundary regions between classes and blur edges between cover types. In DNA a set of distinct cliques is considered as predictors each time a prediction is to be made for a pixel; each time, the clique with the highest degree of homogeneity is chosen. A homogenous *NN* clique consists of pixels with identical cover type. Table 5 lists the set of 21 neighbourhood cliques considered for each prediction.

An index of homogeneity (*IH*) was computed for each clique, and the clique with the highest *IH* score was chosen as NN_i for the current prediction. After exploring several homogeneity indices (see Press (1996) for a list of alternatives) we

Table 5. Neighbourhood cliques in DNA.
A. Numbering convention for pixels in a 5×5 window around a pixel to be classified (pixel 0).

1	2	3	4	5
6	7	8	9	10
11	12	0	13	14
15	16	17	18	19
20	21	22	23	24

B. DNA neighbourhood cliques.

Clique#	Pixel positions	Compass direction
1	4,5,8,9,10,13	North-east
2	9,10,13,14,18,19	East
3	13,17,18,19,23,24	South-east
4	16,17,18,21,22,23	South
5	12,15,16,17,20,21	South-west
6	6,7,11,12,15,16	West
7	1,2,6,7,8,12	North-west
8	2,3,4,7,8,9	North
9	1,2,3,6,7,8,11,12	North-west quadrant
10	3,4,5,8,9,10,13,14	North-east quadrant
11	11,12,15,16,17,20,21,22	South-west quadrant
12	13,14,17,18,19,22,23,24	South-east quadrant
13	2,3,4,m7,8,9,12,13	Northern block
14	6,7,8,11,12,15,16,17	Western block
15	8,9,10,13,14,17,18,19	Eastern block
16	12,13,16,17,18,21,22,23	Southern block
17	7,8,9,12,13,14,17,18,19	Central block
18	3,8,17,22	North to south
19	11,12,13,14	East to west
20	1,7,18,24	North-west to south-east
21	5, 9, 16, 20	North-east to south-west

chose the following index for clique m given pixel i to be classified ($m=1, \ldots, 21$):

$$IH_{m|i} = PDOM_{m|i} \times \Phi\left[-0.5 \times \left(\mathbf{y}_i - \tilde{\mathbf{y}}_{m|i}\right)' \tilde{\boldsymbol{\Sigma}}_{C_i}^{-1} \left(\mathbf{y}_i - \tilde{\mathbf{y}}_{m|i}\right)\right] \times \underset{j \in m}{E}\left[\sup_k \left(\operatorname{Exp}\left[LL_j\right] \middle| C_j = k\right)\right] \Big/ Tr\left(\hat{\boldsymbol{\Sigma}}_{m|i}\right) \times \frac{v_m}{8}, \tag{4}$$

where $PDOM_{m|i}$ is the relative frequency of the most common ML-predicted cover type in clique m, $\tilde{\mathbf{y}}_{m|i}$ is the generalized mean of the (transformed) TM reflectance values of pixels in clique m, $\hat{\boldsymbol{\Sigma}}_{C_i}$ is the generalized covariance matrix of reflectance values for class C_i, Φ is the distribution function of a Gaussian normal distribution, $\underset{j \in m}{E}\left[\sup_k \left(\operatorname{Exp}\left[LL_j\right] \middle| C_j = k\right)\right]$ is the average maximum likelihood of the ML classifications in m, v_m is the number of pixels in m ($\max(v_m) = 8$), $\hat{\boldsymbol{\Sigma}}_{m|i}$ is a $v_m \times v_m$ matrix of average covariance (averaged across TM channels) of reflectance values in m, and Tr is the trace operator (Searle 1982). If more than one clique had the highest IH value, then the final choice of clique was based on a random draw. The highest IH values are obtained for cliques with a single ML class, spectral reflectance values close to those of pixel i, and low variance of reflectance values. The factor v_m in equation (4) corrects for the intrinsic tendency of smaller cliques to have higher IH values than a larger cliques. Estimation of IH values was restricted to cliques entirely inside a PSU. Once the clique with the highest $IH_{m|i}$ value was identified the pixel (i) was classified with the classifiers MATT, SMOOTH, and DOM described next. The clique with the highest $IH_{m|i}$ and augmented by pixel i is henceforth denoted by NN_i^* and the number of pixels in this augmented clique is denoted by v_m^*. We extend the concept of local spatial continuity (Switzer 1980) to NN_i^* and assume accordingly that all pixels in NN_i^* belongs to one class only.

2.7.3.1. *MATT.* Klein and Press (1989) showed that if the class conditional data vectors $\left(\mathbf{y}_{i|C_i=k}\right)$ are multivariate normal in distribution, spatially autocorrelated, and the pixel (i) to be classified is located in a locally continuous clique of pixels of the same class, then the posterior probability of $C_i=k$ is proportional to a matrix T density (Dickey 1967):

$$\text{prob}\left(C_i=k\middle|\mathbf{y}_i,\, NN_i^*\right)\propto\frac{\text{Exp}[LL_i(C=k)]}{\left|\widetilde{\boldsymbol{\Sigma}}_k+\left(\mathbf{Y}_{NN}^*-\widetilde{\mathbf{M}}_k^*\right)\hat{\mathbf{R}}_{NN}^*-\left(\mathbf{Y}_{NN}^*-\widetilde{\mathbf{M}}_k^*\right)'\right|^{v_m^*/2}},\ k=1,\ldots,K, \quad (5)$$

where $\mathbf{Y}_{NN}^*$ is a $5\times v_m^*$ matrix of (transformed) TM reflectance values of pixels in NN_i^*, $\widetilde{\mathbf{M}}_k^*$ is a $5\times v_m^*$ matrix of class k generalized mean reflectance values, and $\hat{\mathbf{R}}_{NN}^*$ is the $v_m^*\times v_m^*$ matrix of the expected average (across TM channels) spatial autocorrelation coefficients between pixels in NN_i^* (see the Appendix for details on estimation of autocorrelation coefficients). The class that maximized equation (5) was chosen as the MATT class prediction for pixel i.

2.7.3.2. *SMOOTH.* In SMOOTH the clique NN_i^* is classified and the class predicted for NN_i^* is assigned to pixel i. For classification of a clique of spatially autocorrelated pixels the generalized mean of their (transformed) reflectance values provides the best predictor (Flygare 1997, Switzer 1980, Mardia 1984, 1987). The generalized mean of $\mathbf{y}$ in NN_i^* was estimated from:

$$\tilde{\mathbf{y}}_{NN}^*=\tfrac{1}{v_m^*}\mathbf{Y}_{NN}^*\times\hat{\mathbf{R}}_{NN}^{*\,-}\times 1_{v_m^*}, \quad (6)$$

where $1_{v_m^*}$ is a $v_m^*\times 1$ vector of ones. Pixel i was then reclassified by the ML classifier in equation (1) with $\tilde{\mathbf{y}}_{NN}^*$ replacing $\mathbf{y}_{\mathrm{i}}$.

2.7.3.3. *DOM.* In DOM pixel i was reclassified to the most frequent ML predicted class in NN_i^*. Ties were broken by random equal probability draws.

2.7.4. *Mixed method (MIX)*

The increase in classification accuracy anticipated from the use of contextual information may be class and method dependent. Hence, a hierarchical set of classification rules that attempt to combine the apparent class-specific strength of different classifiers could be superior (Jia and Richards 1998, Leblanc and Tibshirani 1996). Based on the class-specific classification accuracies achieved during the training phase with the classifiers described so far, we developed a mixed hierarchical classification rule. Let $C_{i|method}$ be the class predicted for pixel i with a given *method* (*method*=(ML, ICM, MICM, MATT, SMOOTH, DOM)). For the New Brunswick site the MIX rules were:

$$C_{i|\mathrm{MIX}}=\begin{cases} wa \text{ if } C_{i|\mathrm{MATT}}=wa \\ tm \text{ if } C_{i|\mathrm{SMOOTH}}=tm \\ C_{i|\mathrm{ICM}} \text{ in all other cases}\end{cases} \quad (7)$$

The expected superiority of MATT to classify *wa* and SMOOTH to perform well for *tm* pixels is exploited in this rule set. For the Saskatchewan site the MIX rules

were:

$$C_{i|\mathrm{MIX}} = \begin{cases} wa \text{ if } C_{i|\mathrm{MATT}} = wa \\ tb \text{ if } C_{i|\mathrm{MATT}} = tb \\ tc \text{ if } C_{i|\mathrm{DOM}} = tc \\ tm \text{ if } C_{i|\mathrm{ML}} = tm \\ C_{i|\mathrm{ICM}} \text{ in all other cases} \end{cases} \tag{8}$$

Here the expected superiority of MATT, DOM, and ML to classify, respectively, *wa, tb*, and *tm* is captured by the rules. In these rule sets the topmost of two rules takes precedence in the case of a rule conflict (e.g. if $C_{i|\mathrm{DOM}} = tc$ and $C_{i|\mathrm{ML}} = tm$ then application of the precedence rule gives $C_{i|\mathrm{MIX}} = tc$).

2.8. *Performance criteria*

The one-stage cluster sampling design of validation data allows statistical inference of the effect of contextual classification on classification accuracy. ML results serve as benchmark and the significance of an improvement is gauged by one-sided t-tests under the null hypothesis of no improvement (Miller 1980). Since the accuracy achieved with the contextual classifiers depends on the ML accuracy, we estimated the expected (conditional) improvement in classification accuracy by a contextual classifier at five pre-specified levels of ML accuracy (0.5, 0.6, …, 0.9). The conditional expectations were obtained by logistic regression analysis (Lloyd 1999) as the logarithm of the odds of a correct contextual classification was strongly and positively correlated ($r > 0.95$) with the corresponding odds of a correct ML classification. Estimates of conditional expectations were bias-corrected (Wiant and Harner 1979).

The impact of spatial homogeneity and spatial autocorrelation of ML classification errors on the performance of contextual classifiers was explored. Spatial homogeneity has been known to impact on the performance of a contextual classifier (Moody and Woodcock 1996, Moody 1998, Foody 1999, Franklin 2000, Hubert-Moy *et al.* 2001). Entropy is an indicator of spatial homogeneity (Moody 1998). An estimate of entropy (E) was obtained for each PSU from the inventory data. As per Gonzales and Woods (1992) the entropy for a PSU was computed as:

$$\hat{E} = -\mathbf{1}_K' \times \hat{\mathbf{P}} \times \log\left(\hat{\mathbf{P}}\right) \times \mathbf{1}_K \tag{9}$$

where $\mathbf{P}$ is a $K \times K$ matrix of relative frequencies of the K^2 ordered pairs of cover type classes observed in left-to-right adjoining pixel pairs in a *PSU*. Random admixtures of small distinct areas have high entropies, and conversely, a mosaic of large homogeneous areas, a low entropy. A minimum is reached if a PSU contains only a single cover type.

Spatial autocorrelation of ML classification errors are expected (Congalton 1988, Masselli *et al.* 1994). The clustering of these errors is likely to impact on the performance of a contextual classifier (Ripley 1985). To quantify this potential relationship we computed a lag one correlation coefficient of (binary) ML errors for each PSU and transformed it to a correlation coefficient of a normal variate (Kedem 1980).

3. Results

The relationships between the accuracy of a contextual and the ML classifier are illustrated in figures 2 and 3. Gains in accuracy due to contextual classification are largest when the ML accuracy is around 70%. Reclassification by ICM achieved the best results on both sites (table 6). Given an ML accuracy between 50% and 80%, ICM improved the accuracy by four to six percentage points ($p<0.05$). For *ML* accuracies below 40% the ICM, gains were negligible and non-significant. PSU-specific ICM gains in accuracy were negatively correlated with the entropy of a PSU ($\hat{\rho}=-0.30$) and the spatial autocorrelation of ML classification errors in a PSU ($\hat{\rho}=-0.28$). The MICM was a close runner up in terms of accuracy gain but with a tendency to be more variable in performance (20% higher standard error of predicted accuracy gain). ICM was, in relative terms slightly less effective in the Saskatchewan (1%) than in the New Brunswick study ($p=0.01$). A stronger positive spatial autocorrelation of ML classification errors in Saskatchewan (0.57 ± 0.02 versus 0.47 ± 0.02 in New Brunswick) probably contributed to this site effect.

SMOOTH results were, as a rule, slightly inferior (1–2%) to those of ICM and MICM ($p>0.05$); yet SMOOTH did better than ICM and MICM in Saskatchewan

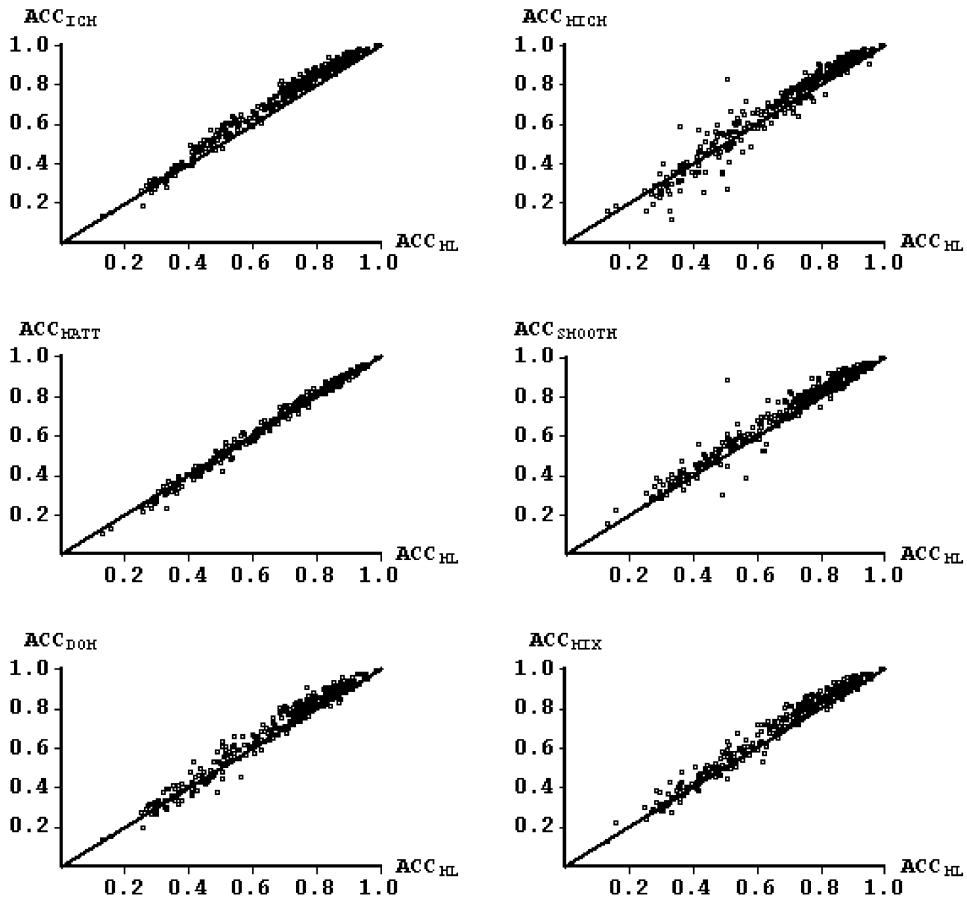

Figure 2. Overall classification accuracy (*ACC*) of the contextual classifiers (*ICM*, *MICM*, *MATT*, *SMOOTH*, *DOM*, and *MIX*) in the New Brunswick study site plotted against the accuracy of the *ML* classification (ACC_{ML}).

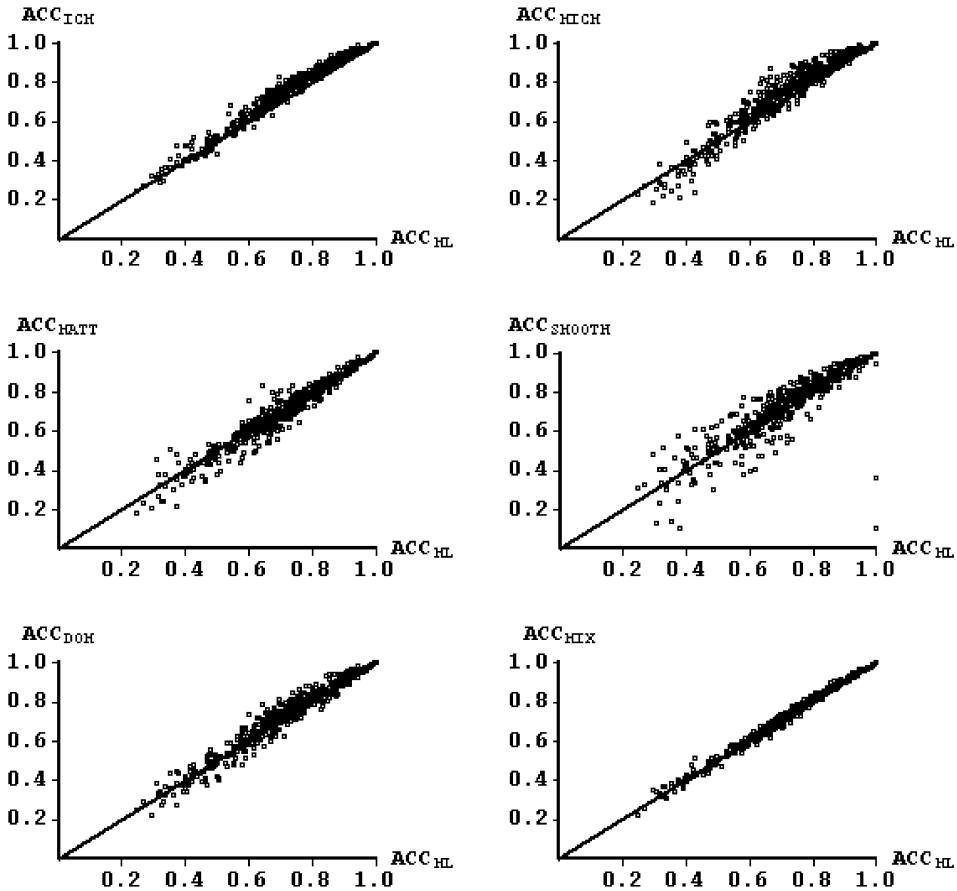

Figure 3. Overall classification accuracy (*ACC*) of the contextual classifiers (*ICM*, *MICM, MATT, SMOOTH, DOM*, and *MIX*) in the Saskatchewan study site plotted against the accuracy of the *ML* classification (ACC_{ML}).

Table 6. Predicted overall accuracy (*ACC*) of six contextual classifiers (ICM, MICM, MATT, SMOOTH, DOM, MIX) at given levels of ML classification accuracy (ACC_{ML}). All table entries are in per cent. Standard errors of predictions are in parentheses. Accuracies significant greater than ACC_{ML} ($p \leqslant 0.05$, *t*-test, Miller (1980) are in bold.

	ACC_{ML}	0.5	0.6	0.7	0.8	0.9
New Brunswick	ICM	**54** (1)	**65** (1)	**76** (1)	**85** (1)	**94** (2)
	MICM	51 (2)	63 (2)	**74** (2)	**84** (2)	93 (3)
	MATT	50 (1)	61 (1)	71 (1)	81 (1)	91 (1)
	SMOOTH	**54** (2)	**64** (2)	**74** (1)	**84** (2)	92 (2)
	DOM	52 (2)	**63** (1)	**74** (1)	**84** (1)	93 (2)
	MIX	**54** (2)	**65** (1)	**75** (1)	**85** (1)	**93** (1)
Saskatchewan	ICM	**54** (2)	**64** (1)	**74** (1)	**84** (1)	**92** (1)
	MICM	53 (2)	**64** (2)	**74** (2)	**84** (2)	93 (2)
	MATT	48 (2)	59 (2)	69 (1)	80 (1)	90 (2)
	SMOOTH	**58** (4)	65 (4)	73 (3)	81 (3)	89 (4)
	DOM	51 (2)	62 (2)	72 (1)	82 (1)	91 (2)
	MIX	51 (1)	61 (1)	71 (1)	81 (1)	91 (1)

when ML accuracy dropped below 60%. DOM produced results at par with those of SMOOTH, and, with one exception in Saskatchewan, there was no significant difference ($p > 0.05$) between the two methods. MATT was only slightly (1%) more effective than ML ($p > 0.05$), and MIX did no better than ICM/MICM in New Brunswick and failed to produce any significant gains in Saskatchewan.

Gains in cover-type-specific classification accuracy by contextual classification were highly variable. Figures 4 and 5 display a select sample of the best results for the purpose of demonstrating the potential of contextual methods. ICM was most consistent across cover types with positive accuracy gains for ML accuracies between 60% (tm) and 80% (el). MICM was generally characterized by negative gains for ML accuracies below 50% and positive (5–12%) but highly variable gains for ML accuracies above 70%. MATT consistently improved classification of water bodies (6–7%) when ML accuracy was between 40% and 90%. However, MATT results for non-water cover types were site-specific, gains would range from 9% to 14% but losses up to 10% were not uncommon either. SMOOTH and DOM either improved or worsened the classification accuracy of a given cover type. MIX was clearly a hybrid with an intermediate performance.

4. Discussion

Contextual classification of Landsat TM images to forest cover types is a logical extension to ML classification. Since the objects to be classified were mostly forest stands with a size many times exceeding the pixel size, the idea that information contained in a neighbourhood clique can improve classification has a sound physical and probabilistic foundation (McLachlan 1991). Most contextual methods consider a TM scene to be a mixture of a 'true' image plus random noise (Pratt 1991) and they aim to reduce the noise (McLachlan 1991). Contextual methods have been most successful when classes have been well separated, as, for example,

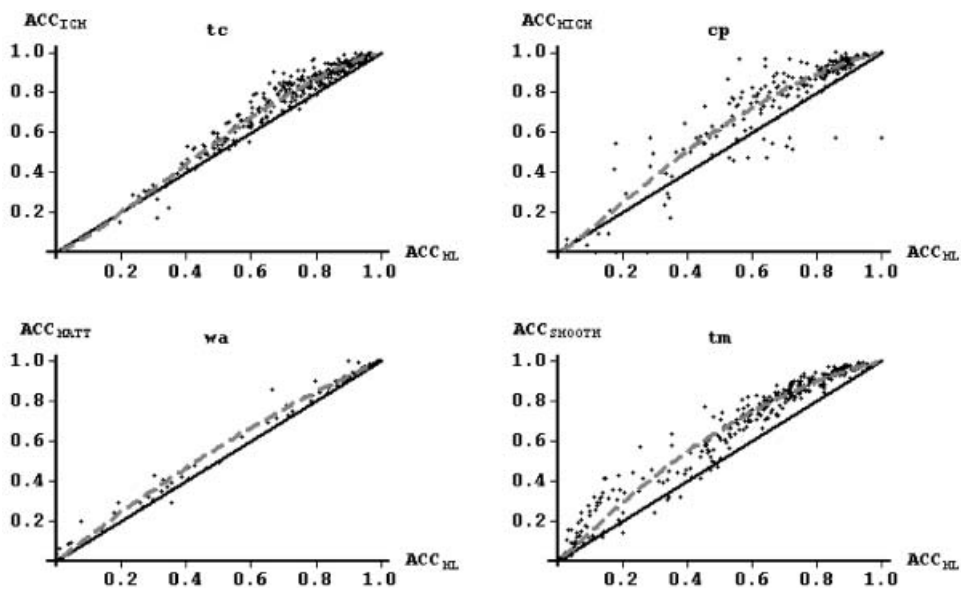

Figure 4. A selection of the best cover-type-specific gains in classification accuracy achieved with the contextual classifiers in New Brunswick. The dashed gray line is trend predicted from a logistic regression. The one-to-one line is provided as a reference.

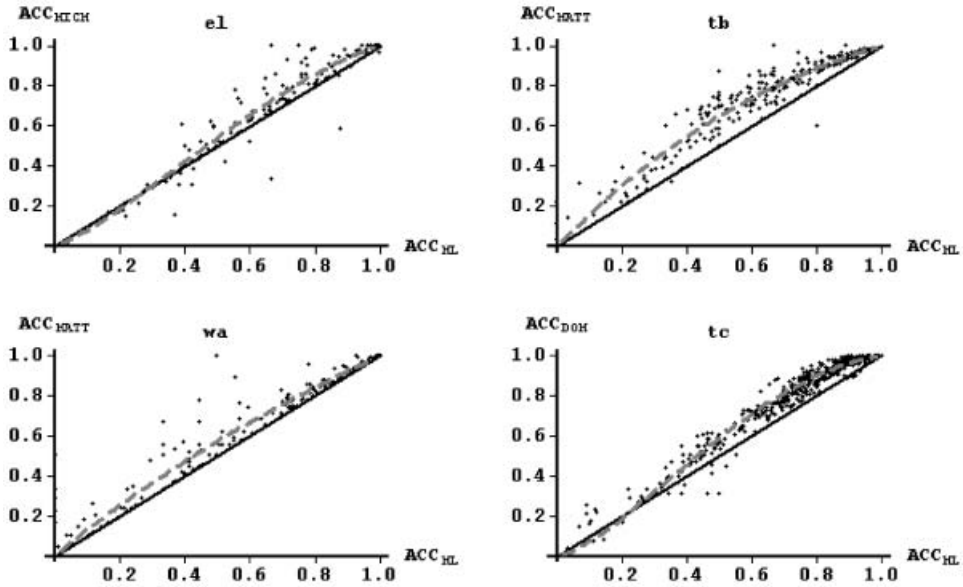

Figure 5. A selection of the best cover-type-specific gains in classification accuracy achieved with the contextual classifiers in Saskatchewan. The dashed gray line is trend predicted from a logistic regression. The one-to-one line is provided as a reference.

in a binary image, or when classes have been defined by non-overlapping univariate class intervals.

Contextual classifiers included in this study represent the basic variants of nearest neighbour methods. ICM and MICM exploit the differential probability of occurrence of various neighbourhood configurations; the remaining methods adopt the concept of local spatial continuity. Working with a fixed neighbourhood configuration as a predictor of cover type creates a bias along boundaries between two classes (Dass and Nair 2003). DNA (Press 1996) was designed to avoid this bias by a data-driven search for the most homogenous neighbourhood clique within a 5×5 search window. However, by choosing a homogeneous clique, the likelihood that the contextual reclassification will be different from the ML classification decreases. This phenomenon is, in our view, the main reason why MATT failed to boost overall accuracy. MATT was developed for black and white images. Accordingly, MATT worked well for the water class where larger water bodies appear as almost black on a brighter background.

The gain in overall accuracy from a contextual reclassification depends on the initial classification accuracy (Klein and Press 1992, Ripley 1985). ML accuracies of about 60% and 80% provide the largest gain. A poor initial accuracy may deteriorate further in a contextual reclassification. The demonstrated improvements of 4–6% compare reasonably well with results published elsewhere (Hubert-Moy *et al.* 2001). For certain cover type classes the gain in accuracy can be substantial (10–12%). Combining classifiers in a way that improves overall classification accuracy has proven difficult (Leblanc and Tibshirani 1996), as confirmed in our study. The performance of the contextual classifiers is expected to improve under conditions with more narrowly defined cover types and more distinct demarcations of forest stands. Many Canadian forest cover types are complex, multi-species,

uneven-aged assemblies of tree species and stand boundaries are often fuzzy (Lowell 1994).

Contextual classification goes beyond the use of spectral values and classification results from neighbouring pixels. Informative priors, such as land suitability class, elevation, topography, land use patterns, and context knowledge can improve classification accuracy (Watson and Wilcock 2001). Bayesian methods (Ripley 1985), fuzzy logic or neural networks (Palubinskas *et al.* 1995) integrates prior knowledge into the classifier. Extending a contextual method to include informative priors is straightforward. Texture methods (Franklin 2000) extend contextual classifiers by exploiting a suite of descriptive statistics for groups of pixels considered as an object. Texture 'signatures' of known spatial objects (Shugart *et al.* 2000) are used for classification of objects created by image segmentation (Hill 1999). Object-oriented classification effectively removes the 'salt and pepper' effects otherwise seen in most classified images (Stuckens *et al.* 2000), but classification accuracy may not improve over and above the accuracy achieved with a per-pixel classifier (Wilson 1996). Our experience with image segmentation (eCognition™, Copyright © 2000, Delphi2 Creative Technologies, Ltd) followed by classification of objects was mixed. Despite obvious improvements in classification of water-bodies, waterways, and roads, the overall accuracy was only one percentage point higher than what was achieved with ML.

Several factors limit the potential of contextual classification methods. They can only improve the classification of pixels that are not surrounded by pixels of the same predicted class. In New Brunswick 47% of the pixels were surrounded by eight pixels of the same inventory (true) class. In Saskatchewan it was 61%. Based on ML classifications the rates increased to 68% and 78%, respectively. Thus, only 32% of the New Brunswick and 22% of the Saskatchewan study site may benefit from a contextual classification. Spatial autocorrelation of classification errors (Wulder and Boots 2001, Congalton 1988) also diminishes the effectiveness of contextual classifiers since the chance that a majority of pixels in a neighbourhood may be misclassified is higher than expected from the overall error rate. When the contextual information is faulty a contextual classifier will lower the classification accuracy. Broadly defined vegetation cover classes characterized by a mixture of multivariate spectral reflectance distributions (Maselli 2001) lowers *a priori* the potential. Forest cover types are cases in point. We consider these limitations as generic and shared by all contextual classifiers. Tree-based classification methods (Li *et al.* 2000) are no exception.

The one-stage cluster sampling design with 2 km × 2 km PSUs and three levels of ML classification accuracy was well suited to quantify the performance of the tested contextual methods. The large number of PSUs allowed us to assess the consistency of the results, and their dependence on ML classification accuracy, spatial autocorrelation of ML errors in the initial classification, and entropy. Cluster sampling of training data is practically and economically attractive (Stellingwerf and Hussin 1997) but adds computational complexities associated with the estimation of means and covariance that are independent of the spatial configuration of the training data (Klein and Press 1989). Ignoring spatial autocorrelation in training data leads to an underestimation of their true variance and, conversely, an inflation of the posterior class conditional probabilities (McLachlan 1991). Generalized variances were in our study from 15% to 60% higher than directly estimated (least squares) variances. Autocorrelation of training data will generally inflate the error rate (see

McLachlan 1991 for further reference) but exceptions have been reported (Dobbertin and Biging 1996).

5. Conclusions

Reclassification of an ML forest cover type classified image with one of the six tested contextual classifiers does not appear to be a very promising option from a cost-benefit perspective. Yet, the potential to improve the accuracy of a single or a few classes significantly appears to be worth exploiting, especially in conjunction with application of 'expert knowledge'. In addition, since the overall classification accuracy in many cases can be increased by 4–6% with the ICM method, we suggest that ICM is applied when a ML classified image doesn't quite meet a predefined quality criteria (Aronoff 1982). In those situations the added costs and time used on reclassification are well spent since the benefit could be acceptance instead of rejection. Although computationally simple, the ICM algorithm adds about 20% to the time needed to classify about 4000 pixels.

Appendix

A.1. *Spatial autocorrelation of reflectance values*

Estimates of the expected autocorrelation of reflectance values between pixels in a neighbourhood clique are needed for estimation of a context-specific class conditional likelihood (e.g. 5) and for generalizing estimates of class-specific means, variances and covariance of reflectance values obtained from cluster sampling (Sherman 1996). The purpose of generalized estimates is to remove a potential effect of the specific spatial configuration of data used for the training of a classifier. Scatter plots of the correlation of reflectance values of training pixels versus the (Manhattan) distance separating them indicated that a simple power function (first-order autoregressive) described the relationship well. For pixels in a regular array the Manhattan distance between two pixels is the sum of the number of rows and the number of columns between them (Upton and Fingleton 1985). Accordingly, the correlation (ρ) of reflectance values between two pixels in locations u and v within a PSU was predicted from the model $\rho(u,v)=\phi^{\delta(u,v)}$, where $\delta(u,v)$ is the Manhattan distance between the two pixels and ϕ is a cover type and TM-channel-specific parameter to be estimated by nonlinear least squares regression (Gallant 1987). Channel-specific and average (across TM channels) expected correlation of reflectance values between pixels in a neighbourhood clique was predicted with these models as were the generalized means and covariance matrices of reflectance (see §A.2).

For the estimation of ϕ we obtained from the training sets robust cover type, TM-channel, and PSU-specific estimates of the correlation (Cressie 1991: equation 2.4.17) between pixels separated by a Manhattan distance of 1, 2, …, 20 pixels. Nonlinear frequency-weighted least squares estimates of ϕ were obtained for each cover type and TM channel. The regression models were in all cases significant ($p<0.01$) with about 80% of the total variation explained by the model. Select estimates of ϕ are listed in table 7. Note the considerable among-PSU and among-SET variation in these estimates, an indication of anisotropy in the autocorrelation process (Cressie 1991, Atkinson and Lewis 2000). In New Brunswick the mixed-wood class had the highest autocorrelation across a given distance, whereas in Saskatchewan the broad-leaved class had the highest. Water 'striping', caused by wave actions, may explain the weak correlation of water pixels. Correlations between pixels of different classes were between one-half and one-third of the

Table 7. Nonlinear least squares estimates of the parameter ϕ in the autocorrelation model of reflectance in the training data (SET1, SET2, SET3) on two sites (NB and SASK). Among PSU standard deviation of ϕ is in parenthesis. Cover type abbreviations are defined in table 2.

	NB			SASK		
	SET1	SET2	SET3	SET1	SET2	SET3
el	0.44 (0.27)	0.33 (0.28)	0.34 (0.22)	0.31 (0.29)	0.40 (0.51)	0.43 (0.24)
sh	0.40 (0.25)	0.56 (0.20)	0.43 (0.36)	0.60 (0.22)	0.56 (0.21)	0.67 (0.20)
tc	0.62 (0.17)	0.61 (0.18)	0.63 (0.15)	0.34 (0.21)	0.43 (0.24)	0.16 (0.23)
tb	0.59 (0.14)	0.59 (0.14)	0.62 (0.16)	0.63 (0.15)	0.66 (0.16)	0.62 (0.20)
tm	0.62 (0.13)	0.65 (0.11)	0.65 (0.17)	0.53 (0.22)	0.57 (0.30)	0.47 (0.25)
wa	0.18 (0.11)	0.10 (0.34)	0.11 (0.46)	0.22 (0.16)	0.14 (0.35)	0.21 (0.35)
cp	0.47 (0.26)	0.48 (0.29)	0.62 (0.28)	n.a.	n.a.	n.a.

corresponding within-class correlations. This indicates that the autocorrelation process across class boundaries is limited to the effects of the point spread function and mixed pixels (Collins and Woodcock 1999).

A.2. *Generalized means and covariance matrices*

The generalized mean reflectance value of a TM channel (say t) of a given cover type class (say i) and PSU (say r) with n_{ir} pixels of the ith class was estimated as a weighted mean with weights proportional to the inverse of the model-based estimate of the expected autocorrelation between pixels of class i in the rth PSU (Klein and Press 1990):

$$\widetilde{\mathbf{y}}_{i(t)r} = \mathbf{1}_{n_{ir}} \times \left(\hat{\mathbf{R}}^{-}_{i(t)r} \times \mathbf{y}_{i(t)r}\right) \times \left(\mathbf{1}_{n_{ir}} \times \hat{\mathbf{R}}^{-}_{i(t)r} \times \mathbf{1}_{n_{ir}}\right)^{-1}, \tag{10}$$

where $\hat{\mathbf{R}}_{i(t)r}$ is a $n_{ir} \times n_{ir}$ matrix of model-based estimates of autocorrelation in channel t reflectance values between class i pixels in the rth PSU. Given the autocorrelation model in §A.1, the computation of the inverse of $\hat{\mathbf{R}}_{i(t)r}$ was straightforward since the inverse is tri-diagonal with only six distinct elements in the three diagonals and 0 elsewhere (Guyon 1995). The generalized covariance matrix (Chilès and Delfiner 1999) of reflectance values for the ith class and rth PSU was obtained from Klein and Press (1990):

$$\widetilde{\mathbf{\Sigma}}_{y_{ir}} = \frac{\sum_{u=1}^{n_{ir}} \sum_{v=1}^{n_{ir}} \left(\mathbf{y}_{ir(u)} - \widetilde{\mathbf{y}}_{ir}\right) \times \left(\mathbf{y}'_{ir(v)} - \widetilde{\mathbf{y}}_{ir}\right)' \times \hat{\hat{\mathbf{R}}}^{-}_{ir}(u, v)}{\sum_{u=1}^{n_{ir}} \sum_{v=1}^{n_{ir}} \hat{\hat{\mathbf{R}}}^{-}_{ir}(u, v)}, \tag{11}$$

where subscripts u and v refer to the uth and vth class i pixel in the rth PSU, respectively, and $\hat{\hat{\mathbf{R}}}^{-}_{ir}(u, v)$ is the row-u column-v element of the inverse to the average (across TM channels) expected correlation matrix of class i pixels in the rth PSU. SET-specific generalized means and covariance matrices were subsequently obtained as frequency weighted averages of PSU-specific estimates.

References

ARONOFF, S., 1982, The map accuracy report: A user's view. *Photogrammetry and Remote Sensing*, **48**, 1309–1312.

ATKINSON, P. M., and LEWIS, P., 2000, Geostatistical classification for remote sensing: an introduction. *Computers and Geosciences*, **26**, 361–371.

BESAG, J., 1986, On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B*, **48**, 259–302.

CHILÈS, J.-P., and DELFINER, P., 1999, *Geostatistics: Modeling spatial uncertainty* (New York: Wiley).

CHOU, P., and BROWN, C. M., 1990, The theory and practice of Bayesian image labeling. *International Journal of Computer Vision*, **4**, 185–210.

COCHRAN, W. G., 1977, *Sampling techniques* (New York: Wiley).

COLLINS, J. B., and WOODCOCK, C. E., 1999, Geostatistical estimation of resolution-dependent variance in remotely sensed images. *Photogrammetry and Remote Sensing*, **65**, 41–50.

CONGALTON, R. G., 1988, Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data. *Photogrammetry and Remote Sensing*, **54**, 587–592.

CRESSIE, N. A. C., 1991, *Statistics for spatial data* (New York: Wiley).

D'AGOSTINO, R. B., BALANGER, A., and D'AGOSTINO JR, R. B., 1990, A suggestion for using powerful and informative tests of normality. *The American Statistician*, **44**, 316–321.

DASS, S. C., and NAIR, V. N., 2003, Edge detection, spatial smoothing, and image reconstruction with partially observed multivariate data. *Journal of the American Statistical Association*, **98**, 77–89.

DICKEY, J. M., 1967, Matric-variate generalizations of the multivariate t distribution and the inverted multivariate t distribution. *Annals of Mathematical Statistics*, **38**, 613–615.

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.

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

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

FRANKLIN, S. E., 2000, Incorporating texture into classification of forest species composition from airborne multispectral images. *International Journal of Remote Sensing*, **21**, 61–79.

GALLANT, A. R., 1987, *Nonlinear statistical methods* (New York: Wiley).

GEMAN, S., and GEMAN, D., 1984, Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE*, **PAMI-6**, 721–741.

GOLDBERG, M., and GOODENOUGH, D. G., 1978, Analysis of a spatial filter for Landsat. *Journal of Applied Photogrammetry and Engineering*, **4**, 25–27.

GONZALES, R. C., and WOODS, R. E., 1992, *Digital Image Processing* (Reading, MS: Addison-Wesley).

GREGOIRE, T. G., 1998, Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, **28**, 1429–1447.

GUYON, X., 1995, *Random fields on a network* (New York: Springer).

HILL, R. A., 1999, Image segmentation for humid tropical forest classification in Landsat TM data. *International Journal of Remote Sensing*, **20**, 1039–1044.

HUBERT-MOY, L., CONTONNEC, A., LEDU, L., CHARDIN, A., and PEREZ, P., 2001, A comparison of parametric classification procedures of remotely sensed data applied on different landscape units. *Remote Sensing of Environment*, **75**, 174–187.

JIA, X. P., and RICHARDS, J. A., 1998, Progressive two-class decision classifier for optimization of class discriminations. *Remote Sensing of Environment*, **63**, 289–297.

KEDEM, B., 1980, *Binary time series* (New York: Marcel Dekker).

KLEIN, R., and PRESS, S. J., 1989, Contextual Bayesian classification of remotely sensed data. *Communication in Statistical Theory and Methods*, **18**, 3177–3202.

KLEIN, R., and PRESS, S. J., 1990, Bayesian contextual classification with neighbours correlated with training data. In *Bayesian and likelihood methods in statistics and econometrics*, edited by S. Geisser, J. S. Hodges, S. J. Press and A. Zellner (North-Holland: Elsevier), pp 337–355.

KLEIN, R., and PRESS, S. J., 1992, Adaptive Bayesian classification of spatial data. *Journal of the American Statistical Association*, **87**, 844–851.

LEBLANC, M., and TIBSHIRANI, R. J., 1996, Combining estimates in regression and classification. *Journal of the American Statistical Association*, **91**, 1641–1650.

LI, K.-C., LUE, H.-H., and CHEN, C.-H., 2000, Interactive tree-structured regression via principal Hessian directions. *Journal of the American Statistical Association*, **95**, 547–560.

LLOYD, C. J., 1999, *Analysis of categorical variables* (New York: John Wiley).

LOWELL, K., 1994, A fuzzy surface cartographic representation for forestry based on Voronoi diagram area stealing. *Canadian Journal of Forest Research*, **24**, 1970–1980.

MAGNUSSEN, S., BOUDEWYN, P., and GILLIS, M. D., 1998, Towards a plot size for Canada's national forest inventory. In *Integrated Tools for Natural Resources Inventories in the 21st Century*, edited by M. Hansen and T. E. Burk. (St. Paul, MI: USDA Forest Service), pp. 116–128.

MARDIA, K. V., 1984, Spatial discrimination and classification maps. *Communication in Statistical Theory and Methods*, **16**, 3749.

MARDIA, K. V., 1987, Correction. *Communication in Statistical Theory and Methods*, **16**, 3749.

MASELLI, F., 2001, Definition of spatially variable spectral endmembers by locally calibrated multivariate regression analysis. *Remote Sensing of Environment*, **75**, 29–38.

MASSELLI, F., CONESE, C., and PETKOV, L., 1994, Use of probability entropy for the estimation and graphical representation of the accuracy of maximum likelihood classifications. *Journal of Photogrammetry and Remote Sensing*, **49**, 13–20.

MCLACHLAN, G. J., 1991, *Discriminant analysis and statistical pattern analysis* (New York: Wiley).

MILLER JR, R. G., 1980, *Simultaneous Statistical Inference*. Second edn (New York: Springer).

MOODY, A., 1998, Using landscape spatial relationships to improve estimates of land-cover area from coarse resolution remote sensing. *Remote Sensing of Environment*, **64**, 202–220.

MOODY, A., and WOODCOCK, C. E., 1996, Calibration-based models for correction of area estimates derived from coarse resolution land-cover data. *Remote Sensing of Environment*, **58**, 225–241.

PALUBINSKAS, G., LUCAS, R. M., FOODY, G. M., and CURRAN, P. J., 1995, An evaluation of fuzzy and texture-based classification approaches for mapping regenerating tropical forest classes from Landsat-TM data. *International Journal of Remote Sensing*, **16**, 747–759.

PRATT, W. K., 1991, *Digital Image Processing* (New York: Wiley).

PRESS, S. J., 1996, The directional neighborhoods approach to contextual classification of images from noisy data. *Journal of the American Statistical Association*, **91**, 1091–1100.

RICHTER, R. A., 1990, Fast atmospheric correction algorithm applied to Landsat TM images. *International Journal of Remote Sensing*, **11**, 159–166.

RIPLEY, B. D., 1985, Statistics, images, and pattern recognition. *Canadian Journal of Statistics*, **14**, 83–111.

SEARLE, S. R., 1982, *Matrix algebra useful for statistics* (New York: Wiley).

SHERMAN, M., 1996, Variance estimation for statistics computed from spatial lattice data. *Journal of the Royal Statistical Society, Series B*, **58**, 509–523.

SHUGART, H. H., BOURGEAU-CHAVEZ, L. L., and KASISCHKE, E. S., 2000, Determination of stand properties in boreal and temperate forests using high-resolution imagery. *Forest Science*, **46**, 478–486.

SLAYMAKER, D. M., JONES, K. M. L., GRIFFIN, C. R., and FINN, J. T., 1996, Mapping deciduous forests in southern New England using aerial videography and hyperclustered multi-temporal Landsat TM imagery. In *Gap analysis. A landscape approach to biodiversity planning*, edited by J. M. Scott, T. H. Tear and F. W. Davis (Bethesda: American Society of Photogrammetry and Remote Sensing), pp. 87–101.

STELLINGWERF, D. A., and HUSSIN, Y. A., 1997, *Measurements and estimation of forest stand parameters using remote sensing* (Utrecht: Ridderprint (VSP)).

STUCKENS, J., COPPIN, P. R., and BAUER, M. E., 2000, Integrating contextual information with per-pixel classification for improved land cover classification. *Remote Sensing of Environment*, **71**, 282–296.

SWITZER, P., 1980, Extensions of linear discriminant analysis for statistical classification of remotely sensed satellite imagery. *Mathematics and Geology*, **14**, 433–444.

UPTON, G. J. G., and FINGLETON, B., 1985, *Spatial data analysis by example. Point patterns and Quantitative data* (Chichester: Wiley).

VAN DEUSEN, P. C., 1995, Modified highest confidence first classification. *Photogrammetric Engineering and Remote Sensing*, **61**, 419–425.

WATSON, N., and WILCOCK, D., 2001, Preclassification as an aid to the improvement of thematic and spatial accuracy in land cover maps derived from satellite imagery. *Remote Sensing of Environment*, **75**, 267–278.

WIANT, H. V., and HARNER, E. J., 1979, Percent bias and standard error in logarithmic regression. *Forest Science*, **25**, 167–168.

WILSON, B. A., 1996, Estimating coniferous forest structure using SAR texture and tone. *Canadian Journal of Remote Sensing*, **22**, 382–389.

WULDER, M., and BOOTS, B., 2001, Local spatial autocorrelation characteristics of Landsat TM imagery of a managed forest area. *Canadian Journal of Remote Sensing*, **27**, 67–75.