Distance to second cluster as a measure of classification confidence

Scott W. Mitchell [a,*], Tarmo K. Remmel [b], Ferenc Csillag [c,1], Michael A. Wulder [d]

[a] *Department of Geography and Environmental Studies, Carleton University, Loeb Building B349, 1125 Colonel By Drive, Ottawa, ON Canada K1S 5B6*
[b] *Department of Geography, York University 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3*
[c] *University of Toronto at Mississauga, Canada*
[d] *Canadian Forest Service (Pacific Forestry Centre), Natural Resources Canada 506 Burnside Road, Victoria, British Columbia, Canada V8Z 1M5*

Abstract

Most image classification algorithms rely on computing the distance between the unique spectral signature of a given pixel and a set of possible clusters within an *n*-dimensional feature space that represents discrete land cover categories. Each scrutinized pixel will ultimately be closest to one of the predefined clusters; different classification algorithms differ in the details of which cluster is considered as closest or most likely, but in general the selected algorithm will label each pixel with the label of the closest cluster. However, pixels expressing virtually identical distances to two or more clusters identify a limitation of this typical classification approach. Conditions for limitations to distance based classification algorithms include when distances are long and the pixel may not clearly belong to any single category, may represent mixed land cover, or can be easily confused spectrally between two or more categories. We propose that retention of the distance to the second closest cluster can shed light on the confidence with which label assignment proceeds and present several examples of how such additional information might enhance accuracy assessments and improve classification confidence. The method was developed with simplicity as a goal, assuming the classification has already been performed, and standard clustering reports are available. Over a test site in central British Columbia, Canada, we illustrate the described technique using classified image data from a nation-wide land cover mapping project. Calculation of multi-spectral Euclidean distances to cluster centroids, standardized by cluster variance, allows comparison of all potential class assignments within a unified framework. The variable distances provide a measure of relative confidence in the actual classification at the level of individual pixels.

1. Introduction

Classification of satellite images based on per-pixel comparisons of spectral signatures is a common and convenient method to produce land cover maps for a wide variety of purposes (Bolstad & Lillesand, 1992; Franklin & Wulder, 2002; Pereira & Setzer, 1993). Distances in multi-spectral space between an observed pixel's multi-spectral signature and all identified multi-spectral clusters (from a clustering algorithm or training procedure) are computed and assessed to find the minimum. Typically, the range of possible class assignments is scrutinized to obtain the closest or, in maximum-likelihood classification (MLC), the maximally likely label being assigned to each pixel (Lillesand, Kiefer & Chipman, 2004). However, confidence in this assignment can vary widely — to take extreme examples, some pixels will have only one highly likely assignment; whereas, others could be more similar spectrally, with the most likely class assigned due to only a marginally higher likelihood (shorter distance). Depending on the constituent land cover elements of a pixel and the potentially small differences among pixel–cluster distances, uncertainty often limits the assignment of a label to a pixel. Thus, it is beneficial to retain the class membership likelihoods for each pixel (or at least the most likely assignments), and use this information to map the confidence in class memberships.

The remote sensing community has recognized the limitations of simple per-pixel classifiers in certain cases and has moved to develop new techniques. Some of the increasingly common techniques involve contextual classifiers (Binaghi

* Corresponding author. Tel.: +1 613 520 2600x2695; fax: +1 613 520 4301.
  E-mail address: Scott_Mitchell@carleton.ca (S.W. Mitchell).
[1] Deceased.

et al., 2003; Magnussen et al., 2004; Tso & Olsen 2005), trainable neural network classifiers (Carpenter et al., 1999; Erbek et al., 2004; Foody et al., 1995; Kavzoglu & Curran, 2003; Olthof et al., 2004), image segmentation (Chen et al., 2006), spectral unmixing (Gilabert et al., 2000; Plaza et al., 2004), and fuzzy algorithms (Andréfouët et al., 2000; Melgani et al., 2000; Wang 1990). The realization that individual pixels are not necessarily pure from a spectral signature perspective requires that classifiers are designed to deal with pixels comprised of mixed land cover (commonly referred to as mixed pixels), especially where optimal spatial resolution data is unavailable (Hsieh et al., 2001). Mixed pixels represent ground areas occupied by more than one land cover category (Foody & Arora, 1996) where two or more labels may be almost equally probable (i.e., the spectral distance between a pixel's spectral signature and multiple cluster centroids is virtually identical), thereby introducing uncertainty to the label assignment at that location. The ability to assess these spectral distances to subordinate cluster centroids and to compare them with the distances of the assigned cluster provides a measure of confidence in the assignment of pixel labels.

Stehman (2001) elucidates that tradeoffs in precision, population representation, assumptions, and accuracy of reference data can directly influence project costs. Although it can be argued that project cost reduction may lead to reduced data quality, Stehman (2001) argues that there are several means for simplifying and reducing the burden of accuracy assessment with little apparent loss in data quality and that such designs deserve consideration. Accuracy assessment (Congalton, 1991; Foody, 2002; Reddy & Reddy, 1996; Stehman, 1997), one of the six key stages in the land cover mapping process (Cihlar, 1999), often receives less than due attention which may reduce user confidence in classification products. User understanding of the classification process is increasingly sophisticated, with map quality often judged not only through accuracy assessment, but through judgement of the rigor and appropriateness of the decisions made at all stages of the classification process. Provision of map users with additional information to aid in their determination whether an image classification meets given application needs is increasingly desired; the user determines from the assembled methods and results whether or not an available classification meets their information needs.

Traditional accuracy assessments are aspatial, and generally rely on single or a limited set of descriptive numbers (Remmel et al., 2005; Stehman, 1997) to summarize the complex spatial patterns of heterogeneous landscapes formed by interacting non-stationary spatial processes. We realize that such generalization is an effective means for conveying generalities, but specific accuracy assessments and construction of confidence bounds is significantly more involved (Csillag et al., 2006). In this research, we propose and demonstrate an approach for portraying confidence in pixel labelling to augment user understanding of map products using information computed during the image clustering or classification training processes. Further, the approach presented in this paper provides a rigorous but conceptually simple tool for implementation in standard land cover classification exercises that provides a measure of

class assignment reliability at each pixel. Our approach was developed from a perspective where the classification in question has already been completed, therefore we aimed for a method which is independent from the classification algorithm employed, with minimal data requirements. To accomplish this, we compare Euclidean distances (in spectral/input data space) to the first and second closest clusters, and provide a means of assessing the label's distinctness or potential confusion with other categories. This requires only the classified map, the input images used in the classification, the means and variances of the clusters or training areas, and a mapping of these clusters to final class labels.

When maximum-likelihood classification is used, the likelihood of each pixel's class assignment is calculated and can be extracted and used to enrich classification results (Foody, 1990; Foody et al., 1992); some studies have gone on to compare each pixel's likelihood to all other possibilities using a calculation of relative entropy (e.g. Dean & Smith, 2003; Maselli et al., 1994) However, while some common software packages provide the option to map these likelihoods (Biehl & Landgrebe, 2002; Foody, 2002; GRASS Development Team, 2006), or they can be estimated and interpolated based on training data (e.g. Steele et al., 1998), the utilization of this information is often limited and typically unincorporated beyond the actual classification.

As with fuzzy classification regimes (e.g. Sarkar, 2000), likelihoods of membership extend from traditional 'crisp' classifications to those that recognize that land cover reflectance does not occur in homogeneous spectral swaths (Metternicht, 2003). The benefit of having membership distances for each class associated with individual pixels serves two main purposes. First, the ratio of first to second spectral distances translates into a labelling reliability measure, reflecting how distinctly a pixel's spectral signature replicates the cluster centroid with which it is associated (assuming equal cluster variability). Second, the distances indicate which land cover categories might be responsible for confusion and simultaneously characterize the degree of that confusion. These benefits can serve a host of applications for the spatial assessment of map reliability (e.g. strengthening sampling design, improving classifier operation), with little additional effort. The approach is further bolstered by a means for assessing the statistical significance of pixel label reliability.

2. Study site

We demonstrate our approach with illustrative data subsets from the Prince George, British Columbia (53°53′N, 122°40′W) region, generated as part of the Earth Observation for Sustainable Development of forests (EOSD) program (Wulder et al., 2003). This data product is extensively documented by Wulder et al. (2003) and Remmel et al. (2005); thus, we will only provide a brief overview here. The nation-wide EOSD mapping efforts focused on the forested ecozones of Canada to produce a land cover map of the forested area of Canada. The project mapped over 450 Landsat-7 ETM+ images to capture the conditions present circa year 2000. EOSD was completed as a partnership project with the Canadian Forest Service and the

1 - Shadow
2 - Cloud
3 - Snow/Ice
4 - Rock/Rubble
5 - Exposed Land
6 - Water
7 - Shrub - Tall
8 - Shrub - Low
9 - Herb
10 - Bryoids
11 - Wetland - Treed
12 - Wetland - Shrub
13 - Wetland - Herb
14 - Coniferous - Dense
15 - Coniferous - Open
16 - Coniferous - Sparse
17 - Broadleaf - Dense
18 - Broadleaf - Open
19 - Broadleaf - Sparse
20 - Mixed Wood - Dense
21 - Mixed Wood - Sparse
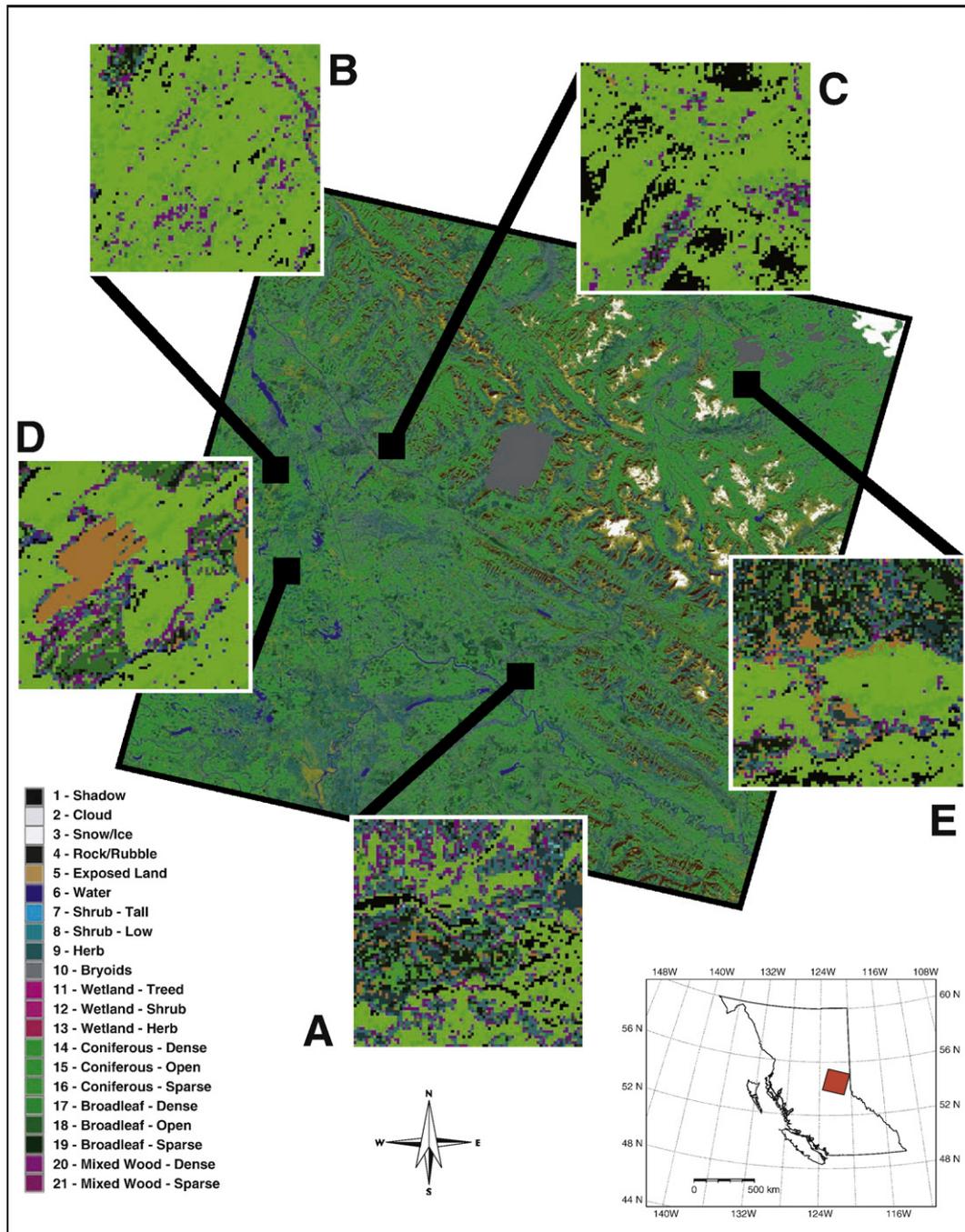
Fig. 1. Locations of five study sites in the Prince George, B.C. region represented by Landsat-7 ETM+ (path 48, row 22) centred at 54°17′13.35″N latitude, 121°8′ 54.70″W longitude, selected to portray differing spatial patterns and compositional structures.

Canadian Space Agency, with the participation of all provinces and territories.[2] For the purposes of this study, the EOSD land cover 2000 product representing the Landsat-7 ETM+ image from Path 48, Row 22, collected on 23 September 2000 is utilized. The classification legend used is based upon the hierarchical classification scheme of the Canadian National Forest Inventory (NFI) (Wulder & Nelson, 2003) which allows

for aggregation of the 22 EOSD thematic classes into sub-groupings of 6, 4, and 3 classes. Remmel et al. (2005) provides an extensive discussion regarding the hierarchical nesting of classes and provides a comparison with the spatially coincident National Forest Inventory (NFI) data (Gillis, 2001; Natural Resources Canada, 1999). The approach presented in this paper was briefly introduced in Remmel et al. (2005) as an ancillary technique that is now rigorously implemented to scrutinize the reliability (confidence) of the EOSD classification on a per-pixel basis. To demonstrate the method here, we concentrate on five

---

[2] EOSD product summary and data access: http://eosd.cfs.nrcan.gc.ca.

(labelled A, B, C, D, and E) 5.76 km^2 (80×80 pixels at 30 m spatial resolution) test regions within the broader study (Fig. 1).

3. Implementation

In addition to the final categorical land cover map resulting from the classification procedure, we are interested in the likelihood values, $P(\mathbf{X}|c)$; that is, the conditional probabilities of each pixel (in vector $\mathbf{X}$) belonging to each class (c) to provide spatial labelling reliability information. This probability is a function of the Mahalanobis distance (Dean & Smith, 2003; Richards, 1993; Richards, 1996), where $\mathbf{V}_c$ is the cluster variance–covariance matrix and m_c is the cluster centroid:

$$D = \log\left(\det|\mathbf{V}_c|\right) + (\mathbf{X} - m_c)^T \cdot \mathbf{V}_c^{-1} \cdot (\mathbf{X} - m_c) \qquad (1)$$

We used this equation as the framework for our logic, but sought a solution requiring less information about the clusters, using typical reports from classification programs as our constraint. Our concern is that the complexity involved with calculating the Mahalanobis distance post-classification may hinder its application, and a simpler method might prove sufficient to provide a quick per-pixel measure of class assignment confidence.

Since the cluster centroids (mean values of each cluster in spectral space) and the main diagonal of the variance–covariance matrix (variance, $\mathbf{V}$, of each cluster in spectral space) are readily available in reports from commonly-used software (and, in our experience, these reports are often retained), we approximate the likelihood using the Euclidean distance of each pixel from its closest cluster, and then account for differing cluster variability by standardizing by the variance ("size") of the respective cluster, using Eq. (2):

$$D = \sqrt{\sum_{i=1}^{n}\left(\frac{r_i - \overline{m}_i}{s_i}\right)^2} \qquad (2)$$

where r_i is the pixel value in the ith band, m_i is the cluster average in the ith band, and s_i is the standard deviation of the cluster in the ith band.

The pixel reliability calculations were programmed using the R open source portable statistical environment (R Development Core Team, 2006). The algorithm requires as input the original image data used for the classification, the cluster statistics provided by the cluster analysis or training procedure, and the mapping of cluster labels to final classes. Our implementation currently utilizes output from PCI Geomatica™ but could be easily adapted to numerous other software packages, requiring only the modification of the import functionality to match the report structure of the clustering software.

The devised function finds the distance from the input value of a given pixel to the cluster leading to the class label assignment, and compares this distance to that of the second closest cluster. This is achieved by calculating the distance from each pixel's multi-spectral value ($n = 80 \times 80 = 6400$ pixels for each sample dataset) to each original cluster centroid (241

clusters in the original image classification), and then finding the two shortest distances from a sorted list, with the added constraint that the second cluster also leads to a different class assignment. The distances and the cluster-class labels are used to produce "as assigned" and "second closest" class assignment (mapping each cluster to the class associated with it in the final classified map) and cluster-distance map layers.

The ratio (d_1/d_2) between the distance to the closest/assigned[3] cluster centroid (d_1) and the distance to the second closest cluster centroid (d_2) along with the magnitude of these distances (at each pixel) provide additional information on the reliability of class label assignment. Since the classification of each pixel is not dependent on its neighbours, the d_1, d_2, and the (d_1/d_2) ratio at each pixel represent independent samples that can be assessed for extreme values. The ratio values theoretically range from nearly 0 (when $d_1 \ll d_2$) to 1 (when $d_1 = d_2$). Given the high number of pixels as samples (i.e., $n > 30$), the Central Limit Theorem states that the distribution of ratio values will be approximately normally distributed. Thus, we convert each ratio to a z-score, conforming to a standard normal distribution with a mean of zero and a standard deviation of 1. In the transformed state, small ratio values will be extremely negative (high confidence in label assignment) while ratio values approaching 1 will be extremely positive (low confidence in label assignment). Using a right-tailed test with $\alpha = 0.05$, the statistically significant, highly positive values are identified using the z-distribution and flagged as having significantly low confidence in their label assignments. These significance maps, in combination with the magnitude of the distances in question, allow evaluation of class labelling confidence across the classified map.

We also compared this Euclidean distance to the Mahalanobis distance (Eq. (1)), using GRASS (GRASS Development Team, 2007) to reconstruct the variance-covariance matrices for each cluster generated by the EOSD method for the entire LANDSAT scene, and then using the built-in mahalanobis() function in R to determine the Mahalanobis distance from each pixel in our first test site (*"Site A"*, discussed below) to each cluster. We compared these distances to the Euclidean distances, and found that while the scatter varied between clusters, the two distance relationships were always positively related, and this relationship was strongest for lower distances. Concentrating on the "closest" and "second closest" clusters identified above, we found that the two measures had good agreement in simple linear fits of the distance to assigned pixel (adjusted $R^2 > 0.6$ for all of our test sites). Furthermore, they agreed on the second candidate cluster for at

---

[3] In the simplest case, the closest cluster and the cluster that leads to the class label assignment would be the same thing. Since we are using the simplified approach of mapping the Euclidean spectral distance (Eq. (2)) as opposed to the likelihood itself, some classification methods (such as MLC) can, in some cases, choose a different class. Therefore, we forced this "first cluster" to always be that which was ultimately used in the final classified map in this study. This occasionally leads to individual pixels where $d_1 > d_2$, so for the algorithm described in this paragraph, these pixels were ignored.
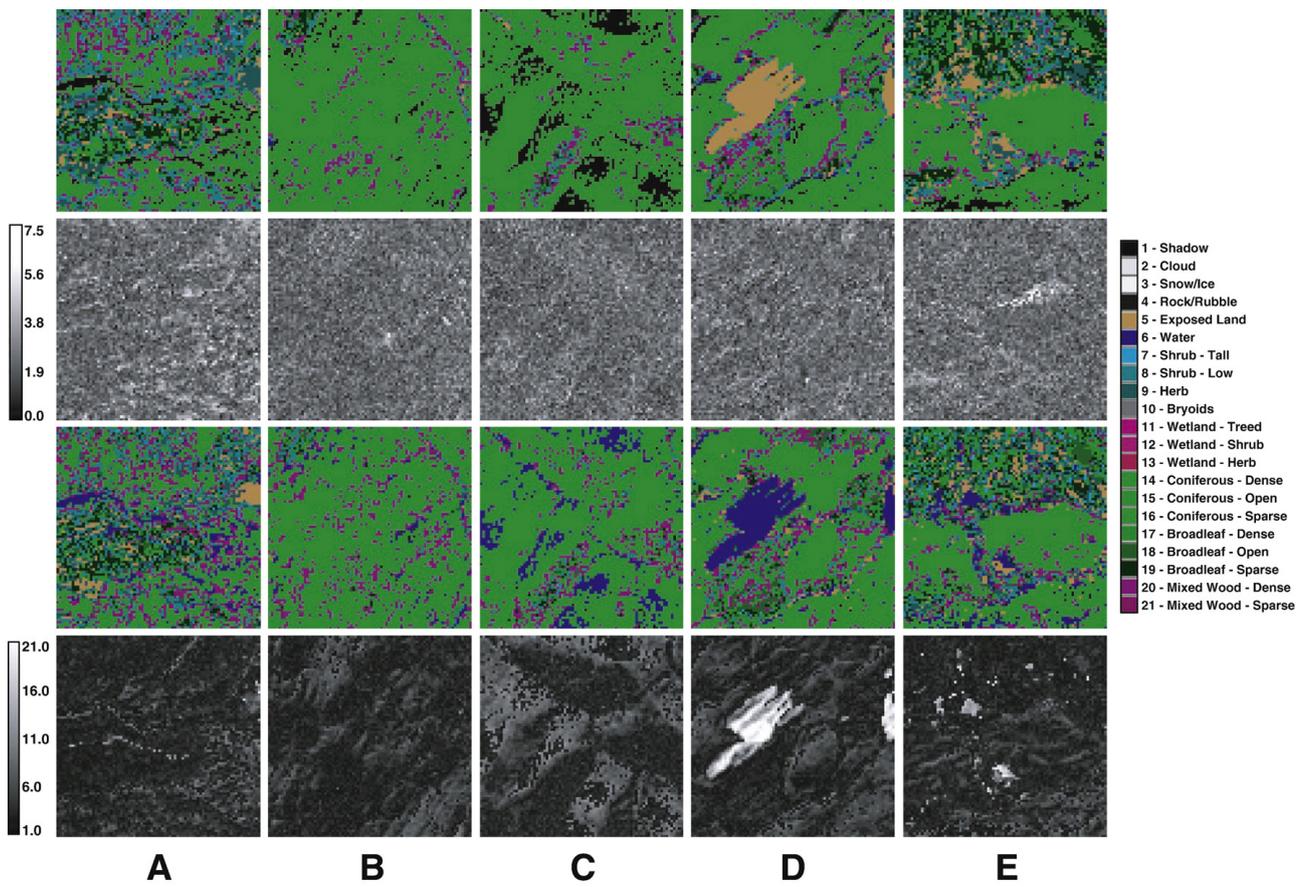
Fig. 2. Assigned classes and standardized cluster distances for the closest and second closest clusters in each study area (each study area makes up a column, labelled A through E). The first and third rows illustrate the land cover classifications for the first and second closest clusters (and therefore class labels) for the five study sites respectively. The second and fourth rows express the standardized distances to the first and second closest clusters.

Table 1
Coincidence matrix between first and second closest class labels assigned to a pixel at study *Site A*

| | | Second class | | | | | | | | | | | | | |
		Shadow	Exposed land	Water	Shrub-tall	Shrub-low	Herb	Confer-dense	Conifer-open	Conifer-sparse	Broadleaf-dense	Broadleaf-open	Broadleaf-sparse	Mixed-dense	Mixed-open	Sum
First class	Shadow	0	4	287	0	0	0	46	0	0	0	0	0	0	0	337
	Exposed land	0	0	61	0	9	81	0	0	19	0	0	1	0	0	171
	Water	0	2	0	0	0	0	4	0	0	0	0	0	0	0	6
	Shrub-tall	0	0	0	0	5	8	0	0	0	1	7	12	0	0	33
	Shrub-low	0	13	0	7	0	352	101	0	27	140	78	145	107	175	1145
	Herb	0	228	0	1	196	0	0	0	34	0	0	61	0	0	520
	Confer-dense	5	0	4	0	90	0	0	1191	0	4	0	0	529	2	1825
	Conifer-open	0	0	0	0	5	0	300	0	0	1	0	3	292	0	601
	Conifer-sparse	0	35	0	0	38	47	0	0	0	0	4	155	0	0	279
	Broadleaf-dense	0	0	0	2	70	0	1	0	0	0	3	5	16	2	99
	Broadleaf-open	0	0	0	3	67	0	0	0	0	9	0	55	12	0	146
	Broadleaf-sparse	0	0	0	9	141	19	0	4	108	11	80	0	23	0	395
	Mixed-dense	0	0	0	0	169	0	226	180	0	82	24	19	0	45	745
	Mixed-open	0	0	0	0	64	0	12	0	0	1	0	0	21	0	98
	Sum	5	282	352	22	854	507	690	1375	188	249	196	456	1000	224	6400

least 90% of the pixels in all sites, and the same areas of low confidence pixels and class confusions were identified. Therefore we elected to continue with Euclidean distance for the rest of this study.

4. Results and discussion

The distances and class assignments for 5 selected test regions (*Sites A* through *E*) are presented in Fig. 2. The

Table 2
Coincidence matrix between first and second closest class labels assigned to a pixel at study *Site B*

| | | Second class | | | | | | | | | | | | | |
		Shadow	Exposed land	Water	Shrub-low	Herb	Confer-dense	Conifer-open	Conifer-sparse	Broadleaf-dense	Broadleaf-open	Broadleaf-sparse	Mixed-dense	Mixed-open	Sum
First class	Shadow	0	1	117	0	0	41	0	0	0	0	0	0	0	159
	Exposed land	0	0	7	1	0	0	0	1	0	0	0	0	0	9
	Water	0	3	0	0	0	5	0	0	0	0	0	0	0	8
	Shrub-low	0	0	0	0	19	18	0	0	12	4	4	3	20	80
	Herb	0	0	0	7	0	0	0	0	0	0	1	0	0	8
	Confer-dense	13	0	14	35	0	0	3688	0	5	0	0	560	0	4315
	Conifer-open	0	0	0	0	0	929	0	0	26	0	12	288	1	1256
	Conifer-sparse	0	0	0	3	3	0	0	0	0	0	1	0	0	7
	Broadleaf-dense	0	0	0	7	0	0	2	0	0	3	3	10	0	25
	Broadleaf-open	0	0	0	8	0	0	0	0	10	0	6	2	1	27
	Broadleaf-sparse	0	0	0	9	0	0	6	12	6	13	0	4	0	50
	Mixed-dense	0	0	0	21	0	148	114	0	116	5	5	0	24	433
	Mixed-open	0	0	0	12	0	1	0	0	0	0	0	10	0	23
	Sum	13	4	138	103	22	1142	3810	13	175	25	32	877	46	6400

Table 3
Coincidence matrix between first and second closest class labels assigned to a pixel at study *Site C*

| | | Second class | | | | | | | | | | | | | |
		Shadow	Exposed land	Water	Shrub-tall	Shrub-low	Confer-dense	Conifer-open	Conifer-sparse	Broadleaf-dense	Broadleaf-open	Broadleaf-sparse	Mixed-dense	Mixed-open	Sum
First class	Shadow	0	0	592	0	0	350	0	0	0	0	0	0	0	942
	Exposed land	0	0	7	0	0	0	0	0	0	0	0	0	0	7
	Water	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Shrub-tall	0	0	0	0	0	0	0	0	0	3	0	0	0	3
	Shrub-low	0	0	0	0	0	4	0	0	24	8	8	31	9	84
	Confer-dense	21	0	19	0	21	0	3679	0	0	0	0	265	0	4005
	Conifer-open	0	0	0	0	3	616	0	0	1	0	10	290	0	920
	Conifer-sparse	0	0	0	0	0	0	0	0	0	0	2	0	0	2
	Broadleaf-dense	0	0	0	0	16	1	1	0	0	1	0	4	1	24
	Broadleaf-open	0	0	0	2	14	0	0	0	5	0	6	12	0	39
	Broadleaf-sparse	0	0	0	0	2	0	3	1	1	7	0	14	0	28
	Mixed-dense	0	0	0	1	47	79	141	0	14	21	21	0	10	334
	Mixed-open	0	0	0	0	7	1	0	0	0	1	0	3	0	12
	Sum	21	0	618	3	110	1051	3824	1	45	41	47	619	20	6400

"Assigned class" map on the top row represents the original EOSD classification. The second row shows the standardized distance to the assigned cluster, and the third and fourth rows respectively provide the class assignment associated with, and spectral distance to, the second closest cluster. In general, pixels with low distances (dark) represent areas with high confidence in the assignment, whereas higher distances indicate potential uncertainty.

Table 4
Coincidence matrix between first and second closest class labels assigned to a pixel at study *Site D*

| | | Second class | | | | | | | | | | | | | | |
		Shadow	Exposed land	Water	Shrub-tall	Shrub-low	Herb	Confer-dense	Conifer-open	Conifer-sparse	Broadleaf-dense	Broadleaf-open	Broadleaf-sparse	Mixed-dense	Mixed-open	Sum
First class	Shadow	0	6	86	0	0	0	46	0	0	0	0	0	0	0	138
	Exposed land	0	0	635	0	1	3	6	0	0	0	0	0	0	0	645
	Water	2	77	0	0	0	0	20	0	0	0	0	0	0	0	99
	Shrub-tall	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2
	Shrub-low	0	2	0	0	0	29	66	0	2	55	21	31	14	55	275
	Herb	0	13	0	0	6	0	0	0	1	0	0	0	0	0	20
	Confer-dense	11	7	14	0	69	0	0	2795	0	9	0	0	352	1	3258
	Conifer-open	0	0	0	0	4	0	362	0	0	44	0	26	175	0	611
	Conifer-sparse	0	5	0	0	1	3	0	0	0	0	0	5	0	0	14
	Broadleaf-dense	0	0	0	4	26	0	5	24	0	0	176	50	156	33	474
	Broadleaf-open	0	0	0	1	34	0	0	0	0	53	0	26	4	2	120
	Broadleaf-sparse	0	0	0	0	12	3	0	13	31	56	15	0	17	0	147
	Mixed-dense	0	0	0	0	63	0	96	84	0	212	22	22	0	41	540
	Mixed-open	0	0	0	0	28	0	4	0	0	7	2	0	16	0	57
	Sum	13	110	735	5	244	38	605	2916	34	438	236	160	734	132	6400

Site A is a highly heterogeneous blend of 14 out of the 21 land cover classes. Site B is dominated by dense coniferous trees, with smaller areas of Mixed Wood — Open. Site C has large changes in relief and aspect leading to patches of shadow. Site D has a large homogeneous patch of exposed land, surrounded by a relatively homogeneous matrix of dense coniferous and broadleaf forest. The northern portions of Site E are similar in nature to Site A, in terms of the mix of classes and spatial heterogeneity, but the southern portions of the area are dominated by larger contiguous patches of dense conifer.

The cluster-distance maps for the 5 sites show interesting differences. Site A has clusters of higher distances to the first cluster throughout the region, with some distinct clustering. Also notable in Site A are roughly linear patterns of high distances to the second cluster, surrounded by patches of low distances. Site B has relatively uniform and low distances to the first cluster, representing good confidence in the cluster memberships across the region and also comparatively low distances to the second cluster across the whole region, but with distinct spatial patches. Sites C and D have relatively uniform, low distances to the first cluster across site extents; although, in Site C the second cluster distances exhibit highly clustered patches of higher distances surrounding the areas that have a lot of shadow. In Site D the spatial pattern is dominated by two distinct patches of high second cluster distances in the areas originally classified as exposed land. Site E has relatively low and uniform distances to the first cluster across much of the image with a cloud of higher distances in the transition zone between the large Coniferous–Dense area and the mixed shrub

and Exposed Land area on its northern extent. The second cluster-distance map illuminates other fairly well defined patches of high distance just south and northwest of the image centre.

In addition to mapping the spatial distribution of classification confidence, more analysis on thematic confusion is possible using coincidence matrices of the first and second class assignments, as presented in Tables 1–5. Interpretation of these tables allows for the identification of classes that are frequently confused, as well as common combinations of classes subject to confusion. Although this is similar to analysis usually performed when validating classified maps against training data or other reference maps, in this case it is comparing every pixel in the classified map with the second most likely class assignment, allowing examination of which categories dominate the first and second classifications, and are therefore most likely to be confused. Observation of these coincidence matrices for multiple test areas within the larger study area indicates class spectral similarity trends, which combined with the spatial information in the distance maps, could be used to highlight problem areas in the classified map. In this case study, the coincidence tables show us that the most common confusion possibilities tend to be between the non-vegetated categories and the density classes within forest types (as also discussed in Wulder et al., 2007). Site C has many pixels potentially confused between Shadow and water, while Site D is at risk of confusion between Exposed Land and Water.

The information presented thus far maps per-pixel relative reliability and identifies possible thematic confusion, but the

Table 5
Coincidence matrix between first and second closest class labels assigned to a pixel at study Site E

|		Second class													
		Shadow	Exposed land	Water	Shrub-tall	Shrub-low	Herb	Confer-dense	Conifer-open	Conifer-sparse	Broadleaf-dense	Broadleaf-open	Broadleaf-sparse	Mixed-dense	Mixed-open	Sun
First class	Shadow	0	13	135	0	0	0	86	0	0	0	0	0	0	0	234
	Exposed land	0	0	293	0	2	148	13	0	16	0	0	3	0	0	475
	Water	3	28	0	0	0	0	12	0	0	0	0	0	0	0	43
	Shrub-tall	0	0	0	0	0	0	0	0	0	10	2	2	0	0	14
	Shrub-low	0	9	0	0	0	182	22	0	22	70	33	143	31	61	573
	Herb	0	247	0	0	105	0	0	0	41	0	0	74	0	0	467
	Confer-dense	28	19	25	0	24	0	0	1923	0	2	0	0	157	2	2180
	Conifer-open	0	0	0	0	8	0	313	0	0	2	0	25	202	0	550
	Conifer-sparse	0	38	0	0	30	49	0	0	0	0	0	107	0	0	224
	Broadleaf-dense	0	0	0	57	21	0	0	6	0	0	164	68	16	0	332
	Broadleaf-open	0	0	0	12	48	0	0	0	3	81	0	52	1	0	197
	Broadleaf-sparse	0	0	0	2	142	102	0	19	245	206	68	0	16	0	800
	Mixed-dense	0	0	0	0	80	0	51	49	0	49	23	16	0	27	295
	Mixed-open	0	0	0	0	12	0	0	0	0	0	0	0	4	0	16
	Sum	31	354	453	71	472	481	497	1997	327	420	290	490	427	90	6400
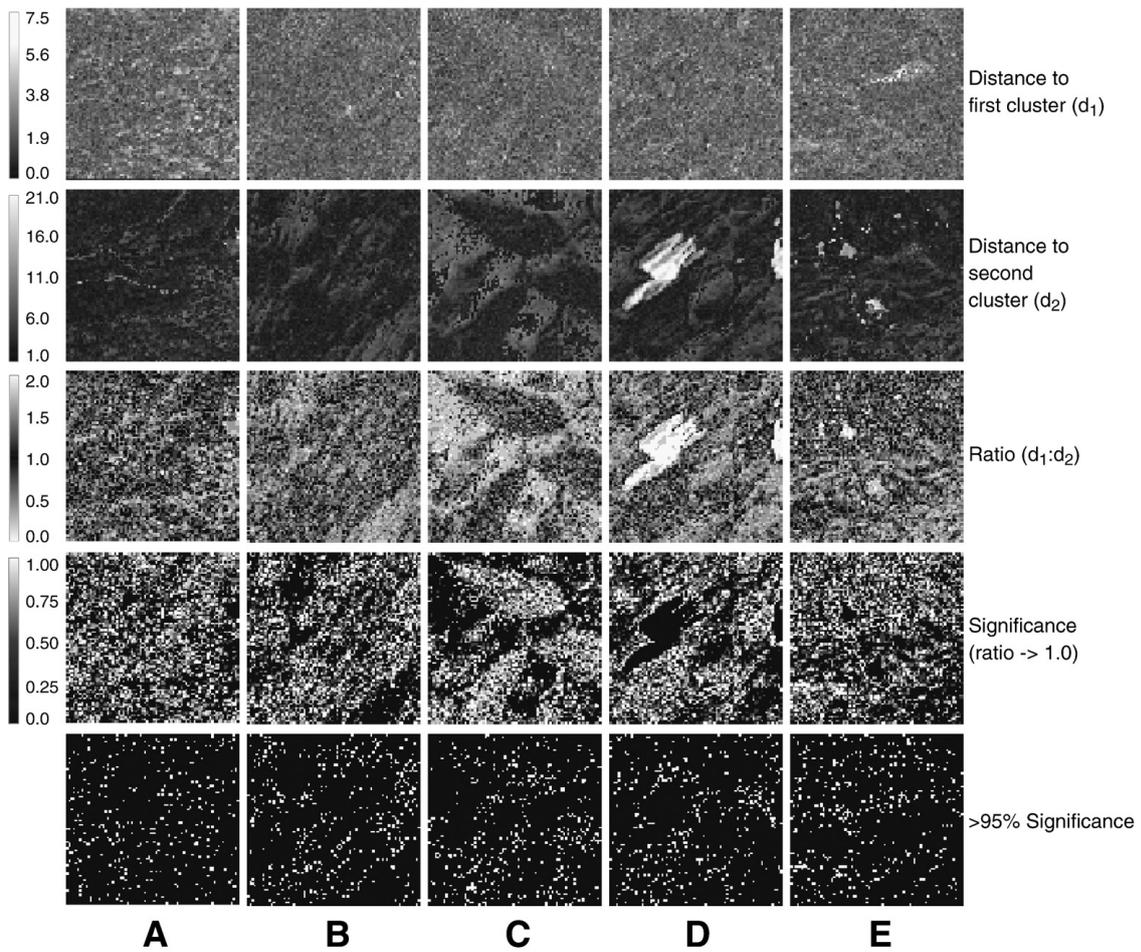
Fig. 3. For each study site (columns A through E), distance to the closest (d_1, first row) and second closest (d_2, second row) clusters, the ratio of the first and second distances (d_1:d_2, third row), the significance value of the z-score of this ratio using the "confusion" tail (i.e., d_1:d_2=1.0; fourth row), and binary maps in which white pixels are those with at least 95% significance (fifth row).

tables alone do not describe how likely it is that the classes are actually confused, and the maps alone do not explain whether or not the distances between the first and second closest clusters are *significantly* different. To address this, distance ratio maps (between first and second spectral classes at each pixel) and *z*-scores for each pixel were calculated (Fig. 3), to further highlight areas of concern by assessing and mapping the degree of potential confusion. It is then prudent to set an alpha level (e.g., $\alpha = 0.05$) for identifying those pixels from the entire population that express statistically significant potential class labelling confusion. The ability to map the outcome has the added benefit of illustrating and allowing the identification of spatial patterns such as clustering or edge effects in the potential confusion.

In *Site A*, the slight differences in cluster distances across the region are amplified when looking at the ratio of the cluster distances, and the significance of these ratios. Although this site remains mixed compared to the other sites, we see some pattern in the ratio and significance images, with generally higher confidence in the more homogeneous portions of the image, compared to areas where the forest is interspersed with shadow and shrubs. The patches of higher distances to second clusters in *Site B* compared to relatively uniform first distances result in corresponding areas of higher classification confidence, showing up as dark patches in the thresholded significance map, indicating no significant confusion. The confused pixels were scattered through the rest of *Site B*. The aspect effect suggested above for *Site C* appears again on the significance map, with high potential confusion in the distinct patches that had low distances to the second cluster, but fairly strong confidence everywhere else. In *Site D* we see a high confidence in the classification of the two large bare land patches, and in a few patches of dense coniferous forest, in the south-east quadrant, but relatively poor confidence elsewhere. In *Site E*, the one vague patch of higher distances to the first cluster does not show up well in the significance map, since there was also a band of higher distances in that same area in the second cluster distances. However, the two main patches of high second cluster distances in areas of low first cluster distances show up as highly confident areas. Comparing these maps to the classifications in Fig. 2, we see these patches are mostly classified as bare rock, and the two with high spatial homogeneity come through clearly as highly confident classifications.

Overall, areas with high degrees of mixed classes in these sample EOSD maps (e.g., some combination of fine extent mixing due to local heterogeneity, ecotones, shadowing) tend to have very similar first and second cluster distances; therefore, the distinction between possible classes is weak. The area immediately south of the large bare patch in *Site D* exemplifies this heterogeneity problem. Similar conditions exist in many of the areas that are dominated by vegetation, but not by a single type; the large potential for confusion between the vegetation classes leads to low overall confidence in these regions, at least at the full level of thematic detail (see Remmel et al., 2005 for effects of thematic aggregation in this data). Areas with more homogenous vegetation, however, such as the larger coniferous regions in *Sites C* and *E*, exhibit generally higher confidence.

The highest confidence areas in all 5 sites are the larger, homogenous blocks of exposed land and rock, which makes sense in terms of spectral separability between these surfaces and the others in these maps.

5. Conclusions

The approach we presented allows for mapping confidence in pixel-based classifications of remotely sensed data. Key aspects of this research include: (1) when Euclidean distances to the first and second most likely classes approach equality, there is a higher chance of misclassification, (2) there is a strong impact of landscape position (e.g., ecotones, gradients) on per-pixel confidence, and (3) these relationships are class-specific (but not site-specific). These findings are so far based on a specific test area, but they follow logically, were demonstrated in a number of detailed test areas, and are expected to hold in other environments. The use of Euclidean distance is a simplification; where the option exists, Mahalanobis distance would be preferred because it accounts for correlation between inputs. In our studies, the difference between these distances was judged to be acceptable for our purposes, but this could change in applications with higher correlations between inputs, or lower differences between cluster distances.

The per-pixel mapping approach is important for applications of land cover data for ecological interpretations (particularly in large-area monitoring studies), because it demonstrates the stochastic nature of observed spatial pattern, with the differences existing in feature space leading to a given class assignment often being slight. Assigned cluster distance, or typicality, mapping is valuable, potentially identifying pixels where insufficient typicality exists to assign a class at all, or to direct further field work (Foody, 1990). We have extended the cluster-distance mapping in this project (Remmel et al., 2005), by determining the distance of each pixel to every possible cluster, and then comparing the closest and second closest cluster assignments and distances in order to evaluate the degree of potential confusion. The methods demonstrated here to evaluate these distances are intended to be used in concert; the distance maps allow visual and quantitative analysis of the spatial pattern of per-pixel confidence, whereas the ratio of distances from alternative classes measures the potential confusion between two alternative classes regardless of their individual distances, and provides an opportunity for hypothesis testing.

Demonstrations show that the relationship between these distances can vary widely within a Landsat scene, and distinct regions of lesser and greater uncertainties are highlighted as a result. This supports the findings of Stehman et al. (2003), where a first or second class is used to validate land cover maps. The approach is also comparable to the "stability maps" produced in Definien's object-oriented classification software, in which the potential confusion between competing memberships for map objects is evaluated. Such information can be used in confidence assessments that build from classified imagery, or to prioritize further investigations (e.g. assessing reliability of existing datasets, guiding further data collection,

etc.). For example, consideration of alternative classes has recently been adapted for use in the accuracy assessment of EOSD products (Wulder et al., 2007). The ratio between first and second cluster distances may be a useful metric in boundary detection, or fuzzy logic applications. Further investigations will include further examination of these possibilities, pursue greater computational efficiency, and develop open source software for incorporating composition and configuration as spatial constraints in per-pixel classification.

Acknowledgements

References

Andréfouët, S., Roux, L., Chancerelle, Y., & Bonneville, A. (2000). A fuzzy-possibilistic scheme of study for objects with indeterminate boundaries: Application to French Polynesian reefscapes. *IEEE Transactions on Geoscience and Remote Sensing, 38,* 257−270.

Biehl, L., & Landgrebe, D. (2002). MultiSpec — A tool for multispectral–hyperspectral image data analysis. *Computers & Geosciences, 28,* 1153−1159.

Binaghi, E., Gallo, I., & Pepe, M. (2003). A cognitive pyramid for contextual classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing, 41,* 2906−2922.

Bolstad, P. V., & Lillesand, T. M. (1992). Improved classification of forest vegetation in northern Wisconsin through a rule-based combination of soils, terrain, and Landsat Thematic Mapper data. *Forest Science, 38,* 5−20.

Carpenter, G. A., Gopal, S., Macomber, S., Martens, S., Woodcock, C. E., & Franklin, J. (1999). A neural network method for efficient vegetation mapping — An introduction. *Remote Sensing of Environment, 70,* 326−338.

Chen, Z., Zhao, Z., Gong, P., & Zeng, B. (2006). A new process for the segmentation of high resolution remote sensing imagery. *International Journal of Remote Sensing, 27,* 4991−5001.

Cihlar, J. (1999). Land cover mapping of large areas from satellites: Status and research priorities. *International Journal of Remote Sensing, 21*(6−7), 1093−1114.

Congalton, R. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment, 37,* 35−46.

Csillag, F., Kabos, S., & Remmel, T. (2006). Confidence in coincidence. *International Journal of Remote Sensing, 27,* 1269−1276.

Dean, A. M., & Smith, G. M. (2003). An evaluation of per-parcel land cover mapping using maximum likelihood class probabilities. *International Journal of Remote Sensing, 24,* 2905−2920.

Erbek, F. S., Ozkan, C., & Taberner, M. (2004). Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. *International Journal of Remote Sensing, 25,* 1733−1748.

Foody, G. M. (1990). Directed ground survey for improved maximum likelihood classification of remotely sensed data. *International Journal of Remote Sensing, 11,* 1935−1940.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment, 80,* 185−201.

Foody, G. M., & Arora, M. K. (1996). Incorporating mixed pixels in the training, allocation and testing stages of supervised classifications. *Pattern Recognition Letters, 17,* 1389−1398.

Foody, G. M., Campbell, N. A., Trodd, N. M., & Wood, T. F. (1992). Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification. *Photogrammetric Engineering & Remote Sensing, 58,* 1335−1341.

Foody, G. M., McCulloch, M. B., & Yates, W. B. (1995). Classification of remotely-sensed data by an artificial neural-network — Issues related to training data characteristics. *Photogrammetric Engineering & Remote Sensing, 61,* 391−401.

Franklin, S. E., & Wulder, M. A. (2002). Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. *Progress in Physical Geography, 26,* 173−205.

Gilabert, M. A., García-Haro, F. J., & Meliá, J. (2000). A mixture modeling approach to estimate vegetation parameters for heterogeneous canopies in remote sensing. *Remote Sensing of Environment, 72,* 328−345.

Gillis, M. D. (2001). Canada's National Forest Inventory (responding to current information needs). *Environmental Monitoring and Assessment, 67,* 121−129.

GRASS Development Team. (2006). *Geographic Resources Analysis Support System (GRASS) Software.* Trento, Italy.

GRASS Development Team. (2007). *Geographic Resources Analysis Support System (GRASS) Software.* Trento, Italy: GNU General Public License.

Hsieh, P. F., Lee, L. C., & Chen, N. Y. (2001). Effect of spatial resolution on classification errors of pure and mixed pixels in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing, 39,* 2657−2663.

Kavzoglu, T., & Curran, P. (2003). The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing, 24,* 4907−4938.

Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2004). Remote sensing and image interpretation. (5th ed.). Wiley (763pp.).

Magnussen, S., Boudewyn, P., & Wulder, M. A. (2004). Contextual classification of Landsat TM images to forest inventory cover types. *International Journal of Remote Sensing, 25,* 2421−2440.

Maselli, F., Conese, C., & Petkov, L. (1994). Use of probability entropy for the estimation and graphical representation of the accuracy of maximum likelihood classifications. *ISPRS Journal of Photogrammetry and Remote Sensing, 49,* 13−20.

Melgani, F., Al Hashemy, B. A. R., & Taha, S. M. R. (2000). An explicit fuzzy supervised classification method for multispectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing, 38,* 287−295.

Metternicht, G. I. (2003). Categorical fuzziness: A comparison between crisp and fuzzy class boundary modelling for mapping salt-affected soils using Landsat TM data and a classification based on anion ratios. *Ecological Modelling, 168,* 371−389.

Natural Resources Canada (1999). *A plot-based National Forest Inventory design for Canada: An interagency partnership project.* In C.F.S. Natural Resources Canada (Ed.).

Olthof, I., King, D. J., & Lautenschlager, R. A. (2004). Mapping deciduous forest ice storm damage using Landsat and environmental data. *Remote Sensing of Environment, 89,* 484−496.

Pereira, M. C., & Setzer, A. W. (1993). Spectral characteristics of fire scars in Landsat-5 TM images of Amazonia. *International Journal of Remote Sensing, 14,* 2061−2078.

Plaza, A., Martínez, P., Pérez, R., & Plaza, J. (2004). A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. *IEEE Transactions on geoscience and remote sensing, 42,* 650−663.

R Development Core Team (2006). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Reddy, M. A., & Reddy, K. M. (1996). Performance analysis of IRS-hands for land use land cover classification system using maximum-likelihood classifier. *International Journal of Remote Sensing, 17,* 2505−2515.

Remmel, T. K., Csillag, F., Mitchell, S., & Wulder, M. A. (2005). Integration of forest inventory and satellite imagery: A Canadian status assessment and research issues. *Forest Ecology and Management, 207,* 405−428.

Richards, J. A. (1993). *Remote sensing digital image analysis — An introduction.* New York: Springer Verlag.

Richards, J. A. (1996). Classifier performance and map accuracy. *Remote Sensing of Environment, 57,* 161−166.

Sarkar, M. (2000). Fuzzy-rough nearest neighbors algorithm. *Systems, Man, and Cybernetics, 2000 IEEE International Conference, vol. 3555,* 3556−3561.

Steele, B. M., Winne, J. C., & Redmond, R. L. (1998). Estimation and mapping of misclassification probabilities for land cover maps. *Remote Sensing of Environment, 66,* 192−202.

Stehman, S. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, *62*, 77−89.

Stehman, S. V. (2001). Statistical rigor and practical utility in thematic map accuracy assessment. *Photogrammetric Engineering & Remote Sensing*, *67*, 727−734.

Stehman, S., Wickham, J. D., Smith, J., & Yang, L. (2003). Thematic accuracy of the 1992 National Land-Cover Data for the Eastern United States: Statistical methodology and regional results. *Remote Sensing of Environment*, *86*, 500−516.

Tso, B., & Olsen, R. C. (2005). A contextual classification scheme based on MRF model with improved parameter estimation and multiscale fuzzy line process. *Remote Sensing of Environment*, *97*, 127−136.

Wang, F. (1990). Fuzzy supervised classification of remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, *28*, 194−201.

Wulder, M., & Nelson, T. (2003). *EOSD Land Cover Classification Legend Report, version 2* (pp. 83). Victoria, BC, Canada: Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre.

Wulder, M. A., White, J. C., Magnussen, S., & McDonald, S. (2007). Validation of a large area land cover product using purpose-acquired airborne video. *Remote Sensing of Environment*, *106*, 480−491.

Wulder, M. A., Dechka, J. A., Gillis, M. A., Luther, J. E., Hall, R. J., Beaudouin, A., et al. (2003). Operational mapping of the land cover of the forested area of Canada with Landsat data: EOSD land cover program. *Forestry Chronicle*, *79*, 1075−1083.