A COMPARISON OF CLUSTERING STRATEGIES FOR UNSUPERVISED CLASSIFICATION

Josef Cihlar[1], Rasim Latifovic[1], Jean Beaubien[2]
1) Canada Centre for Remote Sensing, Ottawa, Canada
2) Canadian Forest Service, Ste-Foy, Quebec

Submitted to:
Canadian Journal of Remote Sensing

Date: October 22, 1999
Revised: February, 2000

Corresponding author:

Josef Cihlar
Canada Centre for Remote Sensing
588 Both Str.
Ottawa, ON
K1A 0Y7
josef.cihlar@ccrs.nrcan.gc.ca

ABSTRACT

A frequently employed image classification strategy is to produce a large number of spectral clusters. However, only a relatively small number of clusters is preferred for labelling, to minimise both analyst effort and the amount of independent information that must be obtained. In this study, we have compared various unsupervised classification algorithms in terms of their ability to define spectrally homogenous and spatially compact clusters. After an initial evaluation of several algorithms, we have focused on two that employ different strategies for locating cluster means: an iterative approach represented by fuzzy K-means (FKM), and a sequential one as used in the Classification by Progressive Generalisation (CPG).

We have also tested a combination of these two techniques. In the initial tests we found that among the iterative algorithms, FKM yielded more spectrally homogenous clusters than others. We then compared initial clusters derived with FKM and CPG in three summer Landsat TM data sets representing different land cover types and patterns from various ecoregions of Canada. After classifying the data to 50 clusters, the average spectral homogeneity was generally (but not always) better for FKM clusters, while average spatial compactness was always higher in CPG. The magnitude of the spectral and spatial differences varied between images. We also found that a combined approach, in which a large number of clusters (80 in this case) is produced by FKM and then merged using CPG, yielded the most consistent clusters among images in terms of both spectral homogeneity and spatial cohesiveness. While the spectral and spatial cluster differences were not large (~10%), they are considered significant in the context of land cover mapping. Although the performance was not compared for a smaller number of clusters, given the principles in the two methods it is expected that the combined approach would perform better for fewer clusters. However, even 50 clusters are sufficiently few to be efficiently labelled using ancillary data. The suitability of this procedure for a particular data set can be readily established using the equations provided.

1.0 INTRODUCTION

Digital classification of multispectral satellite images is commonly used to obtain information on land cover (Cihlar, 2000). In land cover classification, the goal is generally to obtain relatively few classes (~10-30). However, a large number of combinations of spectral values from individual bands is typically found in the images. The aim of classification techniques is thus to reduce the large number of individual combination to a small number of classes. Basically, this approach leads to a classification of pixels, while the objective of land cover classification is an identification of patches/polygons with homogenous land cover characteristics.

The classification algorithms developed since 1970s emphasise efficient use of spectral characteristics of remotely sensed data for two reasons. First, the ability to make accurate spectral measurements is a strong advantage of multispectral sensors onboard satellites or aircraft. Second, digital processing is well suited to analysing spectral properties of individual pixels but poorly able to analyse spatial properties of an image. The initial digital classification algorithms and associated hardware had many limitations, and consequently visual interpretation of enhanced images was a superior method for land cover mapping for a number of years (e.g., Beaubien, 1986). However, at the present time with an automated approach it is possible to extract at least as much as or more information from digital images than an experienced analyst can obtain through visual image interpretation.

Among the various classification approaches, unsupervised image classification methods are designed to make the best possible use of the overall spectral content of an image, with the important condition that these remain as thematically uniform as possible. Two basic strategies have evolved, iterative and sequential. In an iterative procedure such as K-Means or ISODATA (Tou and Gonzalez, 1974), an initial number of desired clusters is selected, and the centroid locations are then moved around until a statistically optimal fit is obtained. In a sequential algorithm such as Classification by Progressive Generalisation (CPG; Cihlar et al., 1998), the large number of spectral combinations is gradually reduced through a series of steps using various proximity measures.

To ensure that no important information is lost at the outset, the initial unsupervised classification is often prepared with a large number of clusters, typically 100-400 for a Landsat TM scene (Homer et al., 1997; Driese et al., 1997; Vogelmann et al., 1998; Cihlar et al., 1998; Beaubien et al., 1999). However, the thematic content cannot be evaluated until the number of clusters is relatively small so that they can be placed into the desired classification legend. Thus, given the large difference between the many spectral clusters and the few land cover classes (~10-30) the question of cluster merging inevitably arises. The spectral merging approach works well up to a point at which the spectrally closest clusters are thematically different, i.e. belong to different classes of the classification legend. If the number of clusters at that point remains large it may even be difficult to ascertain which clusters should not be merged further. This could be because the spatial distributions of the clusters and their relation to land cover are difficult to perceive or because the independent information (ground truth) is not sufficiently detailed. At this stage, the spectral information in the image can be considered to have been used to the full extent. However, there remains the possibility of using spatial similarity criteria. Traditionally, image characteristics such as pattern, texture, shadows etc. provided the basis for land cover mapping through photo interpretation (Colwell, 1960; Rabben, 1960). The interpretation was done to identify land cover polygons, not individual pixels.

In this paper, we evaluate the quality of spectral clusters produced by the sequential and iterative clustering methods. The aim is to obtain spectral clusters that retain as much as possible of the detailed information on vegetation composition (species, density, age, understory) and other surface properties that was present in the original satellite data. We also describe and test a new way of combining spectral and spatial proximity measures in the cluster merging process. This technique can be implemented automatically and is guided only by parameters controlling the classification algorithm. Following an initial comparison of four unsupervised classification procedures, we concentrate on a comparison of two strategies; classification into a pre-specified number of clusters using iterative procedures, and a sequential (uni-directional) merging based on spectral and spatial distance criteria.

2.0 METHODOLOGY

2.1 DATA

We selected three Landsat TM scenes that represent different land cover types and patterns from various ecological regions of Canada (Table 1).

Prior to classification tests, bands 3, 4 and 5 of the image were contrast-stretched (see detailed description in Beaubien et al., 1999). The contrast strech limits were defined using typical scene elements with extreme radiances: dark (water) and bright (broadleaf forest in TM band 4, non-vegetated areas in bands 3, 5). No atmospheric corrections were performed, but this does not affect the results of a relative comparison. The same input data were used for each algorithm.

2.2 CLASSIFICATION ALGORITHMS

The three images were classified using three unsupervised classification algorithms: K-Means and ISODATA (Tou and Gonzales, 1974); fuzzy K-Mean (Bezdek, 1973); and CPGcs (Latifovic et al., 1999); only selected image/algorithm combinations were tested. Table 2 contains the parameters and thresholds for each algorithm.

2.2.1 K-MEANS CLASSIFIER

In K-Means, a sequence of iteration starts with an initial set $\overset{-(0)}{C}$ (Tou and Gonzales, 1974). At each iteration t all pixels $c \in C$ are assigned to one of the clusters $\overset{-(t)}{S}_k$ as defined by nearest neighbour principle. A new centre $\overset{-(t)}{c}_k$ for a cluster is computed as follows:

$$\overset{(t+1)}{c}_j = \frac{1}{N_j} \sum \left(c_i \middle| c_i \in S_j^{(t)} \right).$$

The result of the K-Means clustering is influenced by the number of cluster centres specified, the choice of the initial cluster centres, the order in which the samples are taken, and the geometrical properties of the data.

In this analysis we used PCI's implementation of K-Mean clustering methodology (KCLUS ;PCI, 1999) using the parameters in Table 2. In the PCI implementation, the algorithm locates the initial cluster centres diagonally along the n-dimensional histogram.

2.2.2 FUZZY K-MEANS (FKM)

The fuzzy K-Means algorithm (Bezdek, 1973) is based on the minimisation of the following objective function, with respect to U, a fuzzy K-partition of the data set, and to V, a set of K prototypes:

$$J_q(U,C) = \sum_{j=1}^{N} \sum_{i=1}^{N_{cl,end}} u_{ij}{}^{q} d^2 (X_j, C_i),$$

where X_j is the ***j***-th pixel, C_i is the centroid of the ***i***-th cluster, u_{ij} is the degree of membership of X_j in the i-th cluster, $d(X_j,C_i)$ is the Euclidean distance between X_j and C_i, N is the number of data points, $N_{cl,end}$ is the number of clusters specified (Table 2). The parameter q is the weighting exponent for u_{ij} and controls the "fuzziness" of the resulting cluster. The PCI implementation (FUZCLUS; PCI, 1999) of fuzzy K-mean clustering methodology was used to produce the classified images. In this case, q=2 and the initial cluster centres are located diagonally along the n-dimensional histogram.

2.2.3 ISODATA

The ISODATA algorithm is similar in principle to the K-Means procedure in the sense that cluster centres are iteratively determined (Tou and Gonzales, 1974). However, ISODATA represents a fairly comprehensive set of additional procedures that have been incorporated into an iterative scheme. The ISODATA procedure consists of the following steps: set up processing parameters; distribute the N samples among the present cluster centres; discard sample subsets with fewer than θ_n members; update each cluster centre; compute the average distance of samples in clusters domain; compute the overall average distance of the sample from their respective cluster centre; find the standard deviation vector; find maximum component of standard deviation vector; split clusters; and merge clusters.

2.2.4 CLASSIFICATION BY PROGRESSIVE GENERALIZATION (CPG)

CPG is an unsupervised classification approach which finds means for representative spectral clusters in the data set, assigns every pixel to a cluster, and sequentially combines similar clusters until the remaining clusters can be assigned thematic labels (Cihlar et al., 1998). Latifovic et al. (1999) showed that the sequential cluster merging is more effective if based on spectral proximity constrained by cluster size. This approach is adopted in this paper and is briefly summarised below.

CPG merging procedure

The main decision rule for cluster merging in CPG constrained by cluster size is (Latifovic et al., 1999):

If ($N_{current}>N_{cl,end}$) and ($NP_i<NP_l$) and ($NP_j<NP_l$) and ($SD_{ij}\leq SD_{max}$) then merge. (1)

where $N_{current}$ is the current number of clusters; $N_{cl,end}$ is the number of desired clusters (determined from SD_{ij} table and SD_{max} value); NP_i, NP_j are the sizes of clusters *i* and *j*; SD_{max} is the maximum allowable *SD* for *i*, *j* to merge (Eq. 2); NP_l is threshold cluster size to consider a cluster for merging (Eq. 3); SD_{ij} is the spectral distance of centroids of clusters *i and j* (Eq. 4):

$$SD_{\max} = \sqrt{\sum q_k^{\ 2}} \ , \qquad (2)$$

$$NP_l = NP_a * \frac{N_{cl,st}}{N_{cl,end}} \ , \qquad (3)$$

$$SD_{i,j} = \sqrt{\sum_{k=1}^{N_k} \left(C_{i,k} - C_{j,k}\right)^2} \ , \qquad (4)$$

where:

q_k is number of digital levels per quantised level in the k^{th} spectral dimension (Cihlar *et al.*, 1998);

NP_a is the average cluster size in the input classified image (prior to cluster merging);

$N_{cl,st}$ is the number of clusters in the input classified image;

$C_{i,k}$ is the mean cluster DN value for spectral band *k* of cluster *i*, $i\neq j$;

N_k is the number of spectral bands.

Note that $N_{cl,end}$ and SD_{max} could also be specified by the analyst but the above equations allow automating the process.

Use of spatial distribution in the CPG merging process

Spectral similarity *SD* and spatial coefficient *IC* are used to decide whether clusters *i* and *j* should be merged to create homogenous land cover polygons, with low SD_{ij} and high IC_{ij} values indicating that *i* and *j* are close both spectrally and spatially. Given the many spectral clusters and the frequently mixed pixels in images of land at most spatial resolutions, at least some of the clusters will represent gradients within cover types, i.e. within land cover polygons. These spectral clusters should therefore be merged first.

To consider the spatial attributes of the clusters in merging, spatial adjacency between clusters *i, j*, $SA_{i,j}$, is defined as:

$$SA_{ii} = \frac{NA_{ii}}{NP_i}, \qquad (6a)$$

$$SA_{ij} = \frac{NA_{ij}}{Min(NP_i, NP_j)}, \qquad (6b)$$

where *NP* is the number of pixels in *i* or *j*, and *NA* is the number of times a pixel from the first cluster in the subscript is adjacent to a pixel in the second one. The smaller *NP* value is used in the denominator because of the preference for merging smaller clusters with large ones. *SA* thus refers to the probability of adjacent pixels belonging to the cluster(s) of interest. For one cluster, SA_{ii} can be considered as a measure of clumping and is akin to the contagion index (O’Neill et al, 1988). Spatial coefficient *IC* is then defined as the ratio of intermixing over clumping:

$$IC_{ij} = \frac{2SA_{ij}}{SA_{ii} + SA_{ij}}. \qquad (7)$$

While *SD* and *IC* could be used separately, this creates the problem of their relative importance and weight. Therefore, it is preferable to combine the two measures. One possible way starts with determining *SD* and *IC* of a given cluster in relation to all other clusters. Given the premise that some clusters represent gradients between cover types, clusters within the same cover type should be spatially and spectrally closer compared to those from other cover types. A plot of $SD_{a,j}$ vs. $IC_{a,j}$ for cluster *a* in relation to other clusters *j* should therefore have a consistent shape, decreasing *IC* (Eq.7) as *SD* increases (clusters becoming more different). We have found that an exponential relationship

$$IC = b * SD^d \quad (8)$$

fits best to most clusters, especially for clusters that are spectrally similar (only clusters close than SD_{max} are used to find parameters b and d). An example is shown in Figure 1. The r^2 value for the curves was found to vary between 0.65 and 0.95, for Landsat-derived clusters representing boreal ecosystems. Points on the curve show clusters that show an 'average' spectral and spatial distance from cluster a. Points above (below) the curve represent clusters that are unexpectedly spatially close to (far from) a. Thus, for merging with a, spectrally similar clusters j ($SD_{aj} \leq 3SD_{max}$ used here) which are above the curve should be given first consideration. .

To combine IC with SD, one can find a location on the best-fit curve (Figure 1) which represents the particular combination ($SD_{a,j}$, $IC_{a,j}$). This location will be between two extreme points on the curve corresponding to a) $y=IC_{a,j}$ (i.e., the weight of $SD_{a,j}$ is zero) and b) $x=SD_{a,j}$ (weight of $IC_{a,j}$ is zero). When considering both equally (case c), the point should lie between the two extremes, projected from ($SD_{a,j}$, $IC_{a,j}$). Note that conventional spectral merging without spatial considerations corresponds to case b). Case a) emphasises spatial adjacency; this could lead to contiguous polygons containing spectrally dissimilar cover types. Case c) locates an intermediate position that would be expected given the spatial relationship of a and j in the image and giving equal weight to each; the numerical solution is:

$$SD_{a,j,corr} = SD_{a,j} - bde^{dSD_{a,j,corr}} (be^{dSD_{a,j,corr}} - IC_{a,j}) \quad (9)$$

where $SD_{a,j,corr}$ is the new spectral distance between a and j.

To apply the CPG algorithm, the normal procedure (steps 1-6, Cihlar et al., 1998) was followed that yields a large number of initial clusters (Table 2) after minimum distance classification. Next, the above relations (Eq. 6-9) were employed to correct spectral distances among clusters, and then the clusters were submitted to the decision rule in Eq. 1. Since CPG does not require that the number of clusters be pre-specified, the program was modified to stop merging at the number of clusters desired for comparison with other algorithms.

2.2.5 FKM-CPG

For the purpose of comparison between the merging algorithms, FKM was used to produce classification with 80 clusters, and then Eq. 1 and 4-9 were employed to merge these to 50 clusters.

2.5 ANALYSIS

The analysis was carried out assuming that land cover classes have consistent spectral signatures and secondly, that the resulting land cover polygons should contain as many contiguous pixels as possible, all else being equal. This implies that the spectral clusters should have small within-cluster variability and that individual pixels should have many neighbours belonging to the same cluster. Statistical parameters used to assess cluster homogeneity thus include average standard deviation σ_i (per spectral band k or across the bands, Eq. 10), average within-cluster spectral distance D_i (Eq.11), the spatial coefficient IC (Eq. 7) and the average value of D_i (simple or weighted by cluster size, Eq. 12):

$$\sigma_{i,k} = \sqrt{\frac{1}{NP_i}\sum_{j=1}^{NP_i}\left(x_{j,k} - C_{i,k}\right)^2} \;; \qquad (10)$$

$$D_i = \frac{1}{NP_i}\sum_{j=1}^{NP_i} \sqrt{\sum_{k=1}^{Nc}\left(x_{j,k} - C_{i,k}\right)^2} \;; \qquad (11)$$

$$D_{av} = \frac{\sum_{i=1}^{Nc} NP_i D_i}{\sum_{i=1}^{Nc} NP_i} \;; \qquad (12)$$

where j is the pixel, i is cluster number, k is spectral band, N_k is number of spectral bands, N_c is the number of clusters, and NP_i is the number of pixels in cluster i .

3.0 RESULTS and DISCUSSION

Figure 2 presents a comparison of clusters from the four classification algorithms after pixels in the scene 15-29 were grouped into 50 clusters. Two measures are shown: the average value of the within-cluster standard deviations (Eq.10) in all three bands, and the average of spectral distance of pixels within clusters from the cluster centre D_i. The following observations can be made. First, FKM yielded clusters with the lowest standard deviations and the smallest dispersion about cluster centres . CPG was the closest among the remaining algorithms, and for some measures no appreciable difference with FKM could be discerned. K-Means and ISODATA were consistently worse in terms of the purity of spectral clusters. Consequently, subsequent analysis dealt only with FKM and CPG, both because of the results in Figure 1 and given that these represent different strategies to the generation of spectral clusters.

Figure 3 shows some aspects of the spectral space for FKM and CPG, with circle diameter proportional to cluster size (Figure 3a, 3b) or cluster standard deviation (3c, 3d). Overall, the general characteristics were similar for the two classification methods. The large clusters corresponded to histogram peaks in the two spectral bands (Figure 3a, 3b). However, CPG retained some relatively small clusters away from the peak (3b), while the distribution for FKM was more uniform (3a). The average cluster standard deviation was also very similar for the two classification algorithms (Figure 3c, 3d), with ranges 5.0-20.2 for CPG and 6.0-20.5 for FKM. The more spectrally dispersed clusters were in the margins of the distributions in all spectral bands, although this trend was not uniform as some clusters near the histogram peaks also had higher standard deviations (3c, 3d). The trends shown in Figure 3a-3d were also found in the remaining spectral combinations (not shown).

Figure 4 compares spectral clusters produced by three classification techniques on various images. In all bands, cluster homogeneity varies from scene to scene. The variability among scenes was as high as, or higher than, the variability due to the classification algorithm. Nevertheless, two consistent trends are evident. First, the differences were small when the results of the three scenes are averaged for each spectral band, and neither algorithm is clearly superior. Second, the combined approach employing FKM and the spatial merging component of CPG (Eq. 7) produced the most consistent results overall and the smallest variability among individual images. In other words, in this approach the spectral homogeneity was the most uniform among images and spectral bands. The average values were also the lowest, indicating that the combination of spectral optimisation to a larger number of clusters (FKM), combined with spatial measures in further clustering, helps optimise the clustering process.

The variations of the average within-cluster spectral distance are shown in Figure 5a and 6a- 6c. The variation among scenes was again generally higher than that due to the classification algorithm. Another consistent trend was the substantially lower values for spectral distances weighted by cluster size (Eq. 12), which is due to the disproportionate effect of small clusters. With one exception (scene 15-29), the combined FKM-CPG approach yielded spectrally tighter clusters than either approach separately. This is evident especially when cluster-size weighting is used (Figure 5a). For the same results, the average spatial adjacency index *IC* is shown in Figure 5b. A higher IC value indicates more spatially clumped clusters. As expected, CPG clusters were consistently more cohesive than those produced by FKM. This was also true when comparing the number of distinct patches of pixels belonging to one cluster; for example, the difference was 4.2 % ($2.3x10^5$ patches) for 15-29. In addition, clusters resulting from a combined use of the two methods were similar to those from CPG alone, and in one case (image 15-29) substantially better. Overall, the FKM-CPG approach again gave the best result (Figure 5b). The above results are also supported by an analysis of cluster characteristics based on the Bhattacharya distance measure of separability (as implemented in PCI, 1999).

The differences in IC distributions for various numbers of clusters are illustrated in Figure 5c for image 37-22. In general, CPG tends to produce more spatially compact clusters with higher IC values. However, the difference between the two varied with the number of clusters and was not consistent, especially at the higher IC values.

The overall better performance of the FKM-CPG approach might be expected, for two reasons. First, FKM should in general yield more spectrally pure clusters because the cluster means are optimised for an overall best fit to the distribution of the data; for an initially large number of clusters, this adjustment can be done so that individual pixels are very close to cluster means. In contrast, CPG cluster means are initially established on the basis of the number of pixels with similar spectral values, and are not adjusted through iteration (Cihlar et al., 1998). On the other hand, CPG explicitly takes into account spatial relations between clusters, both in identifying the initial cluster means and during the cluster merging (Latifovic et al., 1999; Eq. 4-9 above). Thus, the FKM-CPG combination should produce clusters that are relatively pure yet spatially compact.

Based on Figure 2, FKM was selected as the representative iterative clustering algorithm for further tests. However, as evident from Figures 3-4, results may vary among scenes. It is therefore possible that K-Means or ISODATA could also provide suitable clusters for the merging process.
The differences among the three approaches were not large, about 10% depending on the measure and the data used. Nevertheless, they may be considered sufficient to justify use of the combined approach. This

is because in land cover mapping, the objective is to identify patches with homogenous land cover characteristics, and patches smaller than a ‘minimum mapping unit’ are ignored. Thus, in amps prepared through digital image analysis, post-processing operations are used to eliminate small patches from the final product, typically by assigning such pixels to neighbouring larger patches. It is thus desirable to minimise the number of such cases. In addition, the CPG approach can also be used to assist the analyst in making further merging or labelling decisions; i.e., the candidates for merging are identified, but the decision is left to the analyst.

Land cover may be characterised by various attributes, not all of which are discernible using spectral reflectance or emission measurements. Classification methods such as neural networks (Carpenter et al., 1997; Bischof and Leonardins, 1998) have been developed to permit use of dissimilar types of data in the classification process. In this case, unsupervised classification can provide one type of input, i.e. spectral characterisation of the land surface cover. Where the spectral content is diagnostic, land cover may be mapped from these data alone. Otherwise, use must be made of other types of information, and of techniques that can combine dissimilar data in consistent decision making rules. In these cases, unsupervised classification can provide the spectral dimension of the decision matrix.

4. CONCLUSIONS

Results of the comparison of the various unsupervised clustering algorithms showed that:

1. The two different clustering strategies (iterative FKM and sequential CPG) produce similar patterns of cluster distribution within the spectral space.
2. For either algorithm, the spectral quality of the resulting clusters tends to vary between scenes and between bands within a scene. While the spectral homogeneity of FKM clusters was generally higher, the differences were fairly small and not entirely consistent. As expected, the spatial cohesiveness was higher for CPG, but the differences were also relatively small at the 50 clusters – stage.
3. The combined FKM-CPG approach produces more consistent results in terms of both spectral and spatial characteristics of the resulting clusters, thus providing the basis for more accurate land cover maps.

Given the variety of land cover types and patterns included in this study, the results may be considered representative for temperate and boreal regions. A more general applicability hinges on the strength of the relationship described by Eq. 8 for a data set of interest. This relationship depends on the spatial resolution of the input data in relation to the spatial and spectral land cover characteristics. Since its strength can be easily determined (Eq. 4, 6, 7, 9), the applicability of the FKM-CPG procedure can be readily established for a given input data set.

5.0 REFERENCES

Beaubien, J. 1986. Visual interpretation of vegetation through digitally enhanced Landsat MSS images. Remote Sensing Reviews 2: 11-43.

Beaubien, J., J. Cihlar, G. Simard, and R. Latifovic. 1999. Land cover from multiple Thematic Mapper scenes using a new enhancement - classification methodology. Journal of Geophysical Research (accepted).

Bischof, H., and A. Leonardis. 1998. Finding optimal neural networks for land use classification. IEEE Transactions on Geoscience and Remote Sensing 36 337-341.

Carpenter, G.A., M.N. Gjaja, S. Gopal, and C.E. Woodcock. 1997. ART neural networks for remote sensing: vegetation classification from Landsat TM and terrain data. IEEE Transactions on Geoscience and Remote Sensing 35: 308-325.

Cihlar, J. 2000. Land cover mappping of large areas from satellites: status and research priorities. International Journal for Remote Sensing (in press).

Cihlar, J., Q. Xiao, J. Beaubien, K. Fung, and R. Latifovic. 1998. Classification by Progressive Generalization: a new automated methodology for remote sensing multichannel data. International Journal for Remote Sensing 19: 2685-2704.

Colwell, R.N. (Ed). 1960. Manual for photographic interpretation. The American Society of Photogrammetry, Washington, D.C. 868p.

Driese, K.L., W.A. Reiners, E.H. Merrill, and K.G. Gerow. 1997. A digital land cover map of Wyoming, USA: a tool for vegetation analysis. Journal of Vegetation Science 8: 133-146.

Homer, C.G., R.D. Ramsey, T.C. Edwards Jr., and A. Falconer. 1997. Landscape cover-type modeling using a multi-scene Thematic Mapper mosaic. Photogrammetric Engineering and Remote Sensing 63: 59-67.

Latifovic, R., J. Cihlar, and J. Beaubien. 1999. Clustering methods for unsupervised classification. Proceedings of the 21st Canadian Remote Sensing Symposium, June 1999, Ottawa, Canada (in press).

O'Neill, R.V., J.R. Krummel, R.H. Gardner, G. Sugihara, B. Jackson, D.L. DeAngelis, B.T. Milne, M.G. Turner, B. Zygmunt, S.W. Christensen, V.H. Dale, and R.L. Graham. 1988. Indices of landscape pattern. Landscape Ecology 1: 153-162.

PCI. 1999. EASI/PACE Software Package, Version 6.3. PCI Inc., Richmond Hill, ON.

Tou, J.T., and R.C. Gonzales. 1974. Pattern recognition principles. Addison-Wesley Publishing Co.

Rabben, E.L. 1960. Fundamentals of photo interpretation. In: Colwell, R.N. (Ed.), Manual of photographic interpretation, The American Society of Photogrammetry, Washington, D.C.: 99-168.

Vogelmann, J.E., T. Sohl, and S.M. Howard. 1998. Regional characterization of land cover using multiple sources of data. Photogrammetric Engineering and Remote Sensing 64: 45-57.

LIST OF TABLES

LIST OF FIGURES

Table 1. Landsat TM scenes used in the tests

Path - Row	Date	Location	Main cover types	Number of pixels	Number of initial clusters in CPG
13-27	14 Aug 1996	Quebec	Forest, wetland, water bodies	$8.36*10^6$	376
15-29	2 Aug 1998	Ontario	Cropland, broadleaf/mixed forest, built-up	$8.85*10^6$	362
37-22	21 Aug 1992	Saskatchewan	Cropland, broadleaf and coniferous forest, wetland, water	$3.49*10^7$	303

Table 2. Control parameters used in the clustering methods

Parameters	**ISODATA**	**K-Means**	**Fuzzy K-Means**	**CPG**
Number of clusters desired	70	70	70	
Maximum number of cluster	75			
Minimum number of cluster	60			
Minimum large seed cluster				0.01
Maximum neglected cluster				0.0002
Minimum cluster size for merging				
Standard deviation	10			
Lumping parameter	1			
Maximum number of pairs of clusters which can be lumped	5			
Number of iterations allowed	20	20	20	
Moving threshold	0.01	0.01	0.01	
Quantisation segment width				10

Figure 1. Plot of spectral distance SD against the spatial coefficient IC for one cluster. Only clusters sufficiently spectrally close to the cluster of interest are included in the computation.

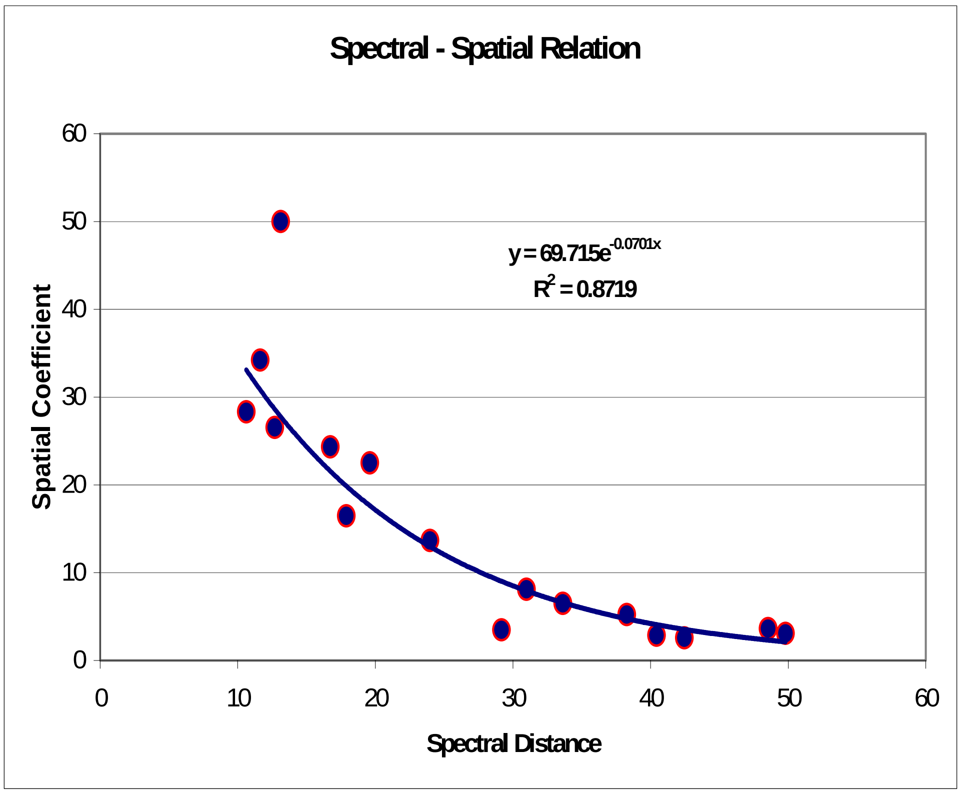


Figure 2. Comparison of five cluster measures for three classification methods. Each data point represents an average value for the 50 clusters. Std and (w) stands for 'standard deviation' and 'weighted by cluster size', respectively.

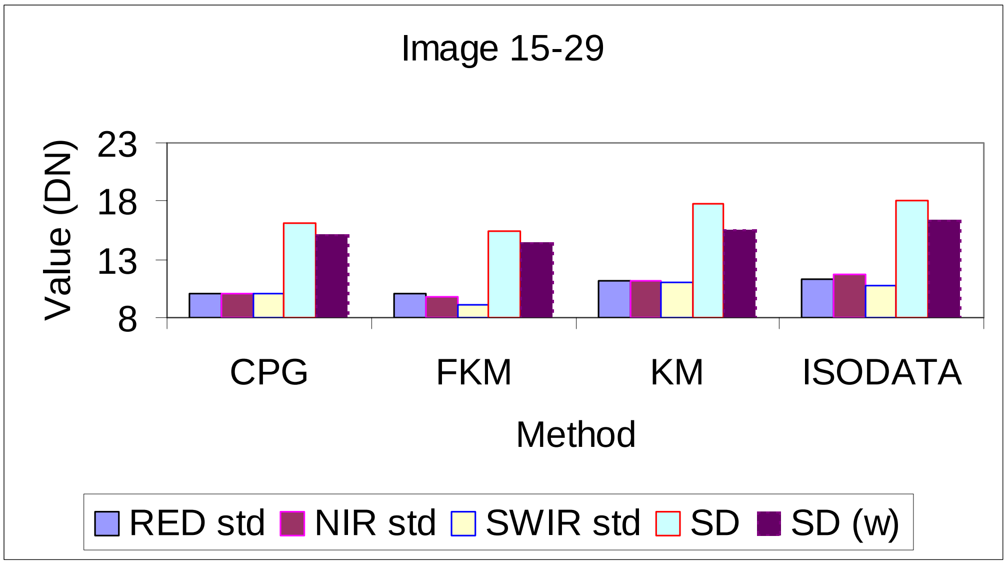
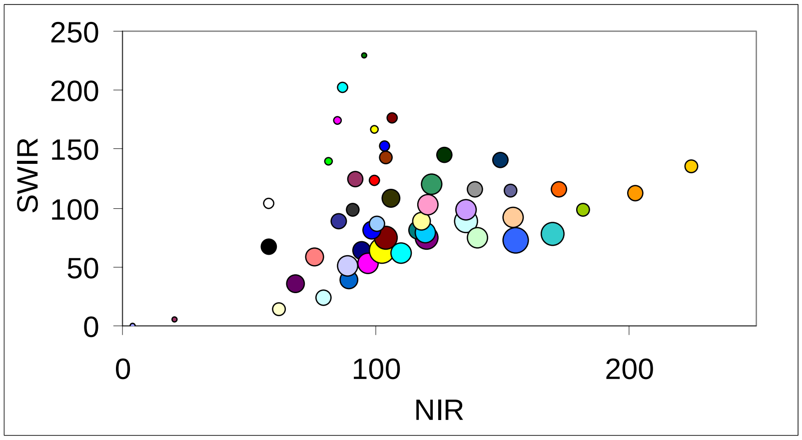
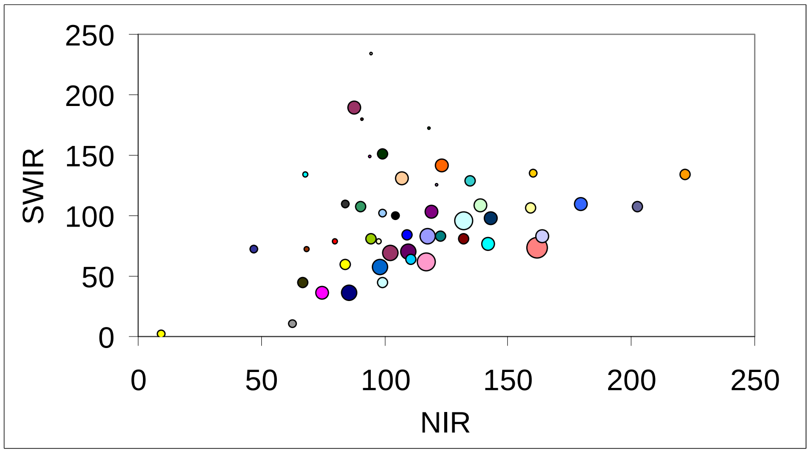


Figure 3. Mean spectral values for clusters obtained by FKM (Figure 3a, 3c) and CPG (3b, 3d). The diameter of the circles is proportional to cluster size (3a, 3b) or average cluster standard deviation (3c, 3d).





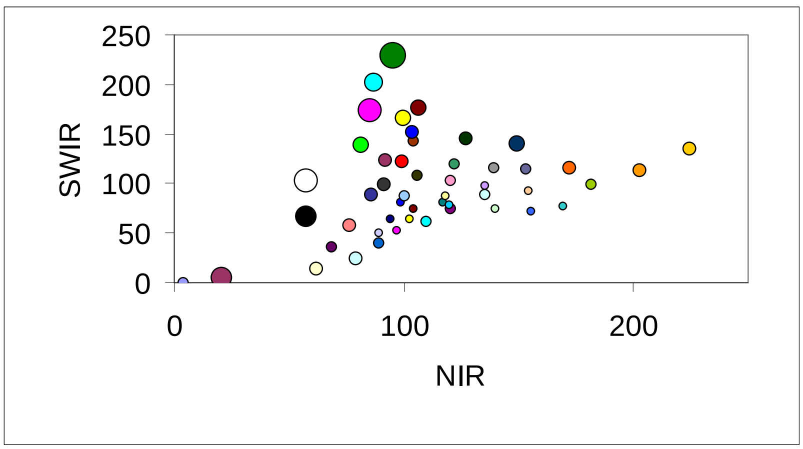
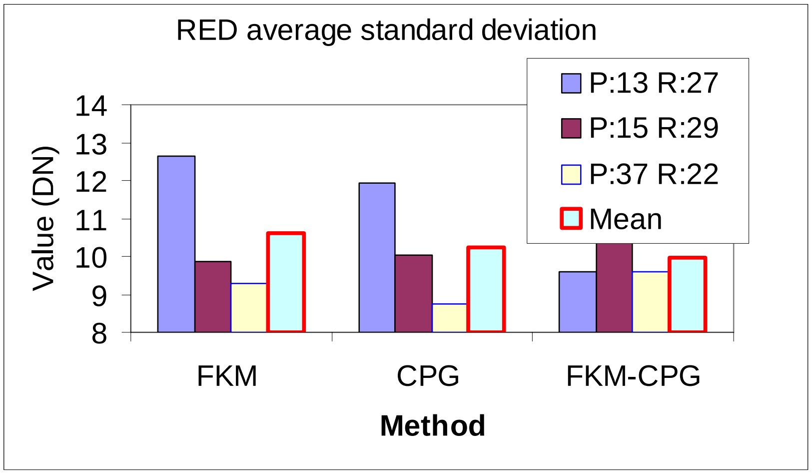
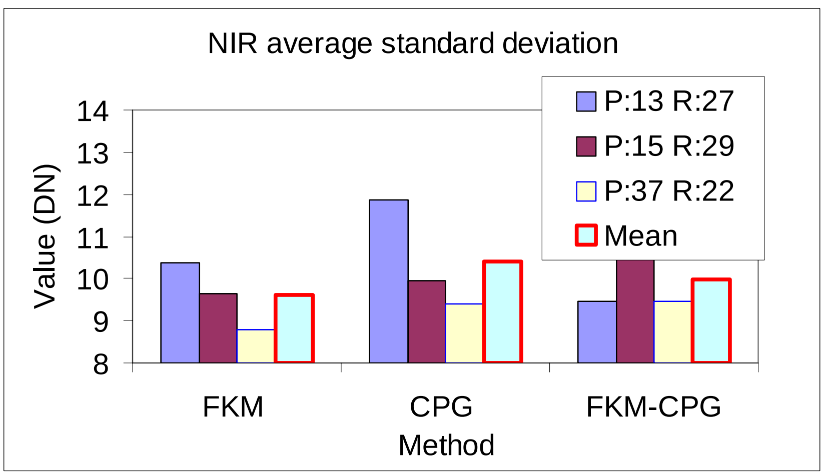


Figure 4. Average standard deviation spectral clusters obtained in different ways for three scenes. Each point represents an average of standard deviations for 50 clusters. The spectral bands are red (TM band 3, Figure 4a), near infrared (TM 4, Figure 4b), and shortwave infrared (TM5, Figure 4c).





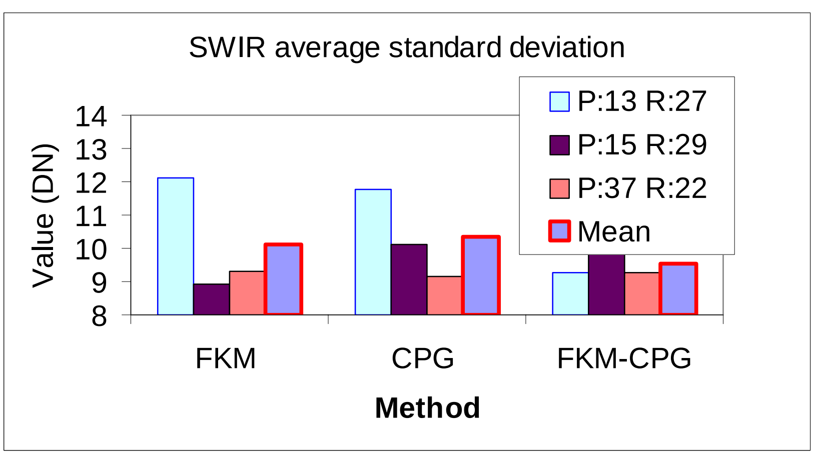
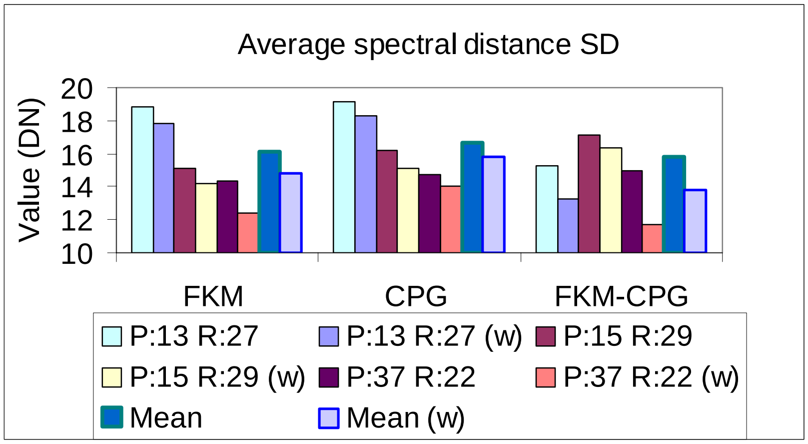
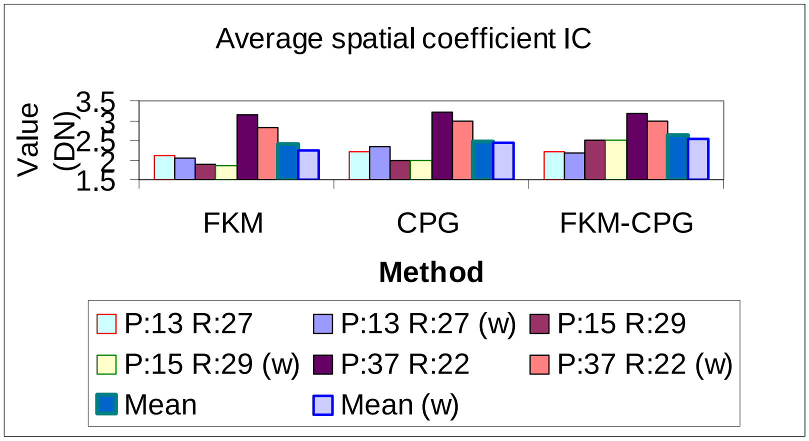


Figure 5. Spectral and spatial measures of relations among clusters obtained by different images and classification methods, for 50 clusters. Figure 5a: Average spectral distance. Figure 5b: Average spatial coefficient. Figure 5c. Histogram of spatial coefficient values. Symbol (w) stands for 'weighted by cluster size'.





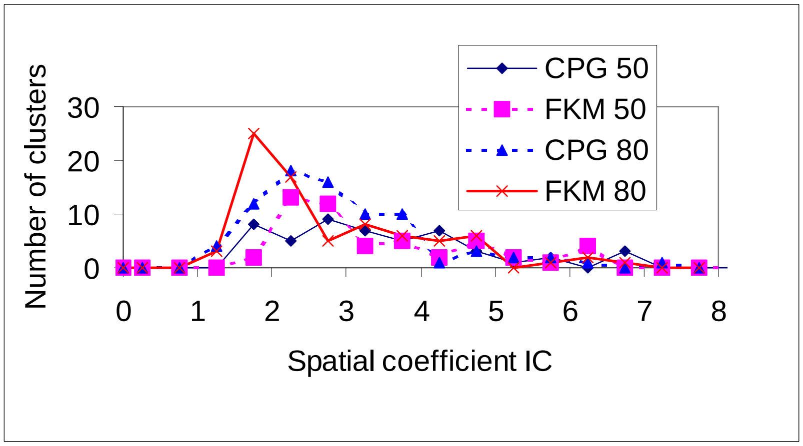


Figure 6. Average spectral distance within clusters D_{av} obtained with different methods for three images. Each point represents an average of spectral distance for all pixels within a cluster.