LANDSCAPE PATTERN AND PER-PIXEL CLASSIFICATION PROBABILITIES

Scott W. Mitchell,

Department of Geography and Environmental Studies, Carleton University,
Loeb Building B349, 1125 Colonel By Drive, Ottawa, ON Canada K1S 5B6
Tel: +1-613-520-2600 x2695
Fax: +1-613-520-4301
E-mail: Scott_Mitchell@carleton.ca

Tarmo K. Remmel, Ferko Csillag,

Department of Geography
University of Toronto at Mississauga, Mississauga

and Mike Wulder

Pacific Forest Research Centre
Canadian Forest Service, Victoria

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. Typically, a range of possible class assignments is considered, and the most likely assignment is chosen for each pixel. But confidence in this assignment can vary widely – to take extreme examples, some pixels will have only one highly likely assignment whereas others could be relatively close to multiple classes, with the most likely class assigned due to marginally higher probabilities. Alternatively, one could "keep" the probabilities of each class assignment, or at least the most likely assignments, and use this information to map the confidence in class membership.

We use the Prince George, British Columbia study area of the Earth Observation for Sustainable Development of forests (EOSD) data product, which covers approximately 130,000 km^2 (6 Landsat scenes) with 146417077 pixels assigned to 22 categories (Wulder et al. 2003). The classification scheme consists of five, hierarchically nested levels (with water/land separation at the coarsest and three forest density classes at the finest level). For about 1% of the area the National Forest Inventory digital data is available (comprising 11049 polygons in 324 2km-by-2km plots arranged on a regular grid). The NFI uses standardized data collection protocols, attributes, and reporting abilities to provide consistent data through which national statistics and reports are compiled to represent Canada's forests (Wulder et al. 2004). We have been comparing these datasets to evaluate their potential for mutual support, starting with a traditional accuracy assessment using coincidence matrices (Remmel et al. 2005).

2. IMPLEMENTATION

Instead of using the land cover map as the final product of the classification, we utilize the likelihood values: P(**X**|c), i.e. the conditional probabilities of each pixel (in vector **X**) belonging to each class (c). The probability can be can be estimated using the Mahalanobis Distance (Richards 1993):

$$P(X\,|\,c) = \log(\det\ |V_c|) + (X - m_c)^T * V_c^{-1} * (X - m_c) \qquad (1)$$

where $\mathbf{V}_c$ is the cluster variance-covariance matrix and m_c is the cluster centroid.

Since the cluster centroids (mean values of each cluster in spectral space) and the main diagonal of the variance covariance matrix (variance, v, of each cluster in spectral space) are readily available as a report from clustering algorithms in commonly used software, we approximate the likelihood using the distance of each pixel from its closest cluster, standardized by the "size" (variance) of the cluster:

$$\text{Standardized Distance} = \sqrt{\left(\frac{X-m_1}{v_1}\right)^2 + \ldots + \left(\frac{X-m_n}{v_n}\right)^2} \qquad (2)$$

A module for the open source GRASS GIS has been developed, which calculates and maps these values given input maps to the classification process, the cluster and class assignments, and a simple report on the means and variances of the clusters.

The software optionally finds the likelihood of the *second* most likely class. This is accomplished by calculating the distance of each pixel from each cluster, and sorting these distances to find the closest cluster which does not lead to the same class assignment in the classified map. Both the size of this distance and the difference of these values from the classification standardized distance give additional information on the confidence we should have in the class assignment.

3. INITIAL RESULTS

This poster demonstrates the application of the software described above to the EOSD classification, and compares per-pixel classification likelihoods to mismatches in the EOSD and NFI data.

Figure 1 shows the map of standardized distance across the entire study area, based on the EOSD classification. At this scale, it is difficult to see the higher distance pixels, but they are spread throughout the map. Contractual limitations in our data use agreement limit our ability to present maps comparing the classified imagery and reference NFI data, but Figure 2 provides an overview of the relative agreement between these maps across the study area (Remmel et al, 2005). One can qualitatively assess that areas with poor agreement often have high uncertainties as mapped by these cluster distances, which may help explain the confusion in those areas. For example, in the southeastern portion of the study area, the agreement map (Figure 2) shows many examples of relatively low agreement, and this area also has many pixels coloured yellow and orange in the distance map (Figure 1). Likewise, there are clusters of uncertainty in the northeast, and we see particularly low relative agreements in the lower levels of aggregation in that area in Figure 1. The difference between these two corners of the maps in terms of agreement as a function of aggregation suggests that there are different factors affecting the classification reliability. Related work (Remmel et al, 2005) has demonstrated that classification disagreement is often highest along transition areas between distinct physical features or ecotones, and the cluster distances also tend to increase along these edges. Distances can also be mapped according to class assignments, highlighting differences in the strength of association with each class.

We have recently added the ability to calculate the distance to the "second closest" cluster, i.e. the cluster that would have resulted in the second most likely class assignment. An example map of this measure compared to the first cluster is presented in Figure 3, zoomed in to a subregion for illustrative purposes, using the same colour assignment as the overall map in Figure 1. The "Assigned class" map on the left represents the standardized distance discussed above. Pixels with low distances (dark) represent areas with high confidence in the assignment, whereas higher distances indicate uncertainty. However, we gain even more information comparing this map to the distances to the closest alternative assignment; if the distance to a cluster that would have resulted in an alternative classification is not much farther than the closest cluster, there is less confidence than when the second closest possibility is far away. On the map of "second class" distances, the left half of the selected area has relatively low distances, while pixels on the right side of the map are generally further from the second closest cluster. This means that, in general, there is more uncertainty in the class assignments on the left side of the map, except for those individual pixels that also have high "Assigned class" distances. We are currently experimenting with possible overlays of thes e maps to provide an index of classification uncertainty, using differences or ratios of the first and second distances; for example, Figure 4 presents the simple pixel differences of the maps in Figure 3.

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

Figure 1 - Standardized distance of each pixel from the cluster used in EOSD classification

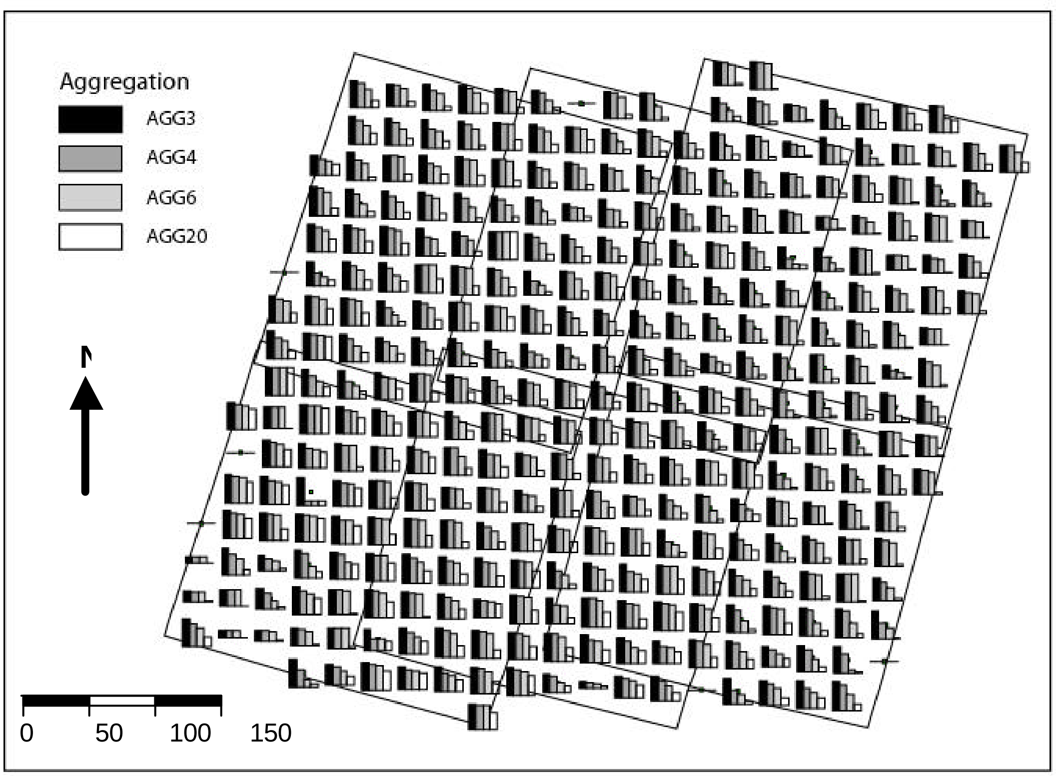

Figure 2 – Agreement between EOSD and NFI classification across 4 levels of aggregation of the classes. The height of the bars represents the level of agreement, and the number in the aggregation label indicates the number of classes in the aggregated classification.

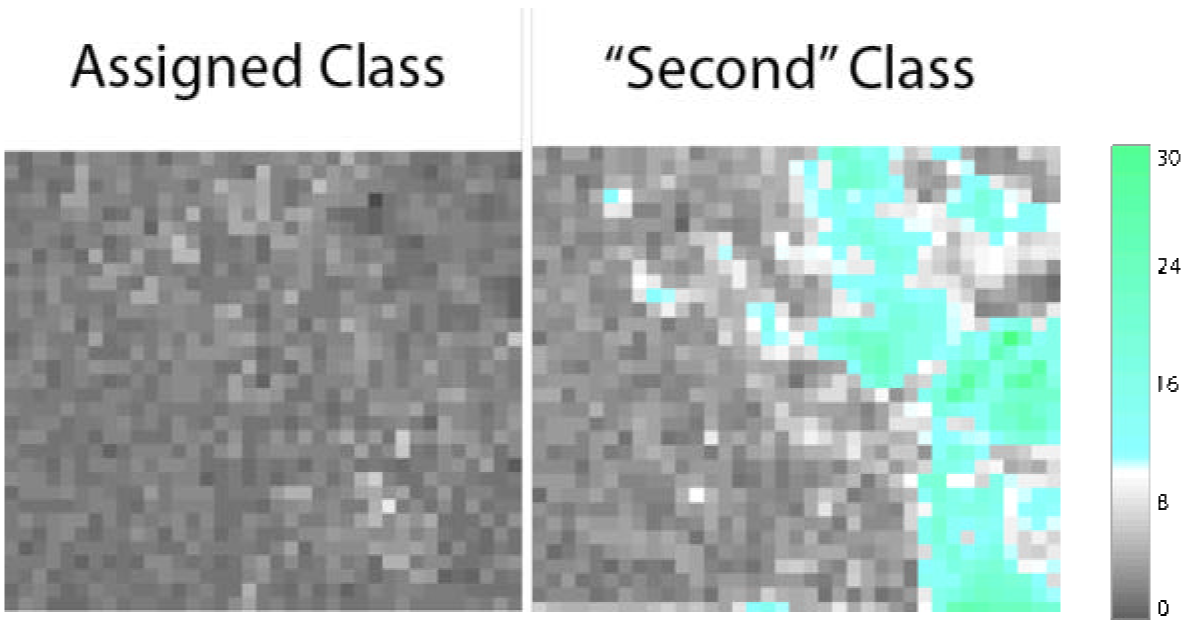

Figure 3 – Standardized distance to the closest (leading to assigned class) and "second closest" (leading to a different class) clusters.

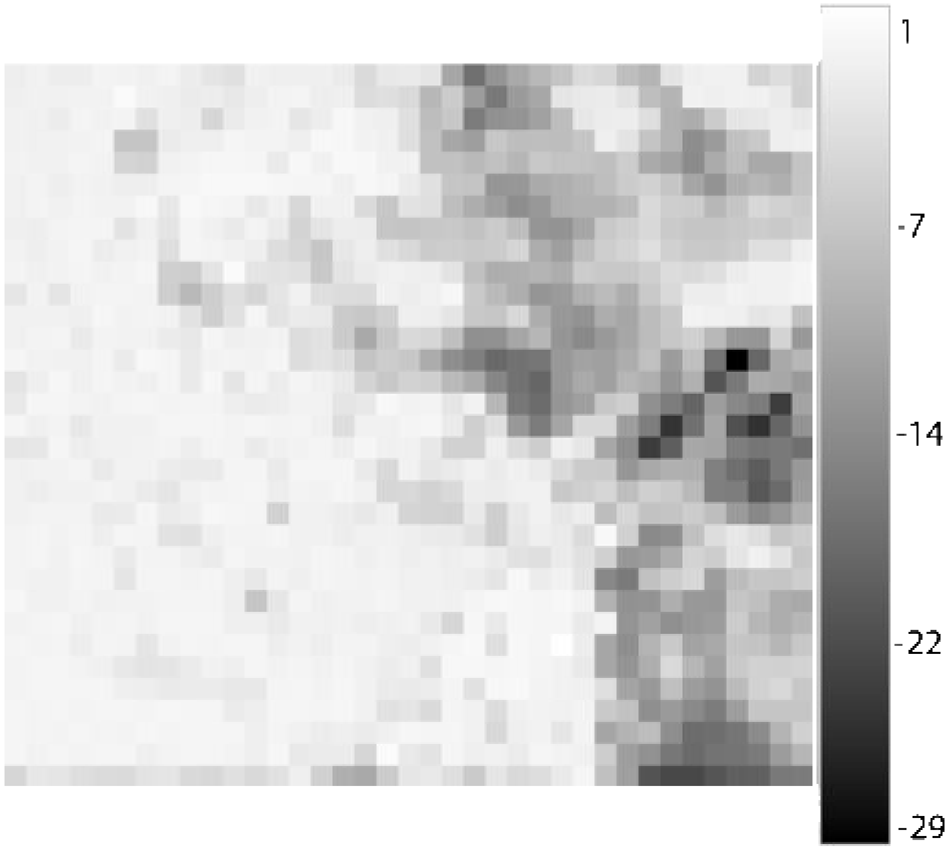

Figure 4 – Difference between the first (assigned class) and second closest clusters for the same area used in Figure 3. White pixels indicate areas with little difference between the assigned class and the closest alternative, therefore the confidence in that classification is low.

4. CONCLUSIONS

The methods we present allow us to map confidence in pixel-based classification. In initial testing of the software on a forest classification application, we have found that (1) there is a relationship between misclassification rates and similarity in likelihood values, (2) there is a strong impact of landscape position (e.g., ecotones, gradients) on likelihood values, and (3) these relationships are class-specific (but not site-specific). These findings are important for applications of land cover data for ecological interpretations (particularly in large-area monitoring studies), because they demonstrate the stochastic nature of observed spatial pattern. We are investigating further model and open source software development for incorporating composition and configuration as spatial constraints in per-pixel classification.

5. REFERENCES

Remmel, T.K., F. Csillag, S.W. Mitchell, and M. Wulder, 2005, Integration of remotely sensed and forest inventory data for national monitoring: Canadian status assessment and research issues. Forest and Ecosystem Management, 207(3), 405-428.

Wulder, M.A., Dechka, J.A., Gillis, M.A., Luther, J.E., Hall, R.J., Beaudoin, A., Franklin, S.E. (2003) Operational mapping of the land cover of the forested area of Canada with Landsat data: EOSD land cover program, The Forestry Chronicle 79, 1075-1083.

Wulder, M.A., Kurz, W.A., Gillis, M. (2004) National level forest monitoring and modeling in Canada. Progress in Planning, 61, 365-381.