CLUSTERING OF POINT PATTERNS DERIVED FROM LIDAR CANOPY HEIGHT DATA

Steen Magnussen[a] and Michael A. Wulder[a]

[a] Natural Resources Canada, Canadian Forest Service, Pacific Forestry Center, Victoria BC V8Z 1M5, Canada - steen.magnussen@nrcan.gc.ca

KEY WORDS: Forestry, LIDAR, Statistics, Classification, Model, Pattern, Segmentation, Method

ABSTRACT:

High intensity canopy height LIDAR data affords model-based estimation of tree locations. The analysis of spatial point patterns is a natural extension of this modeling capability. Identification of within-stand clusters (features) of trees deviating significantly in height from those of surrounding trees (clutter) is important for inventory and forest management purposes. We demonstrate a nonparametric profile likelihood estimation of spatial clusters using Voronoï tesselation with and without prior smoothing via a morphological closure operation on the sets of Voronoï cells considered as solutions to the clustering problem. Smoothing yields not only a more regular outline of clusters but also appears to perform significantly better when there is more than one cluster in the point pattern. Two examples derived from LIDAR canopy data collected above Douglas-fir-dominated stands on Vancouver Island (British Columbia, Canada) illustrate practical applications. The morphological closure of the Voronoï cells prior to computing the likelihood provides more appealing results with potential for practical application in forestry.

1. INTRODUCTION

Forest stands often contain pockets of trees with a height that is distinctly different from that of surrounding trees. These pockets, or clusters as we shall call them, may be due to species differences, age differences, site factors, or the outcome of differing silvicultural treatments. Clusters of trees with similar height characteristics occupying a large enough area may influence significantly the growth, yield, or value of a stand. The identification and spatial delineation of such clusters either during or after a forest stand inventory will improve the precision of stand-level growth and yield predictions and the accuracy of stand-level inventories. There is a plethora of techniques available for identifying and delineating spatial clusters; excellent reviews of approaches and methods can be found in, for example, Ripley (1985), Diggle (1983), Upton and Fingleton (1985), Lawson and Denison (2002), and Cho (2004). The method we have chosen to illustrate appears promising for practical applications in forestry; it is free of the assumptions that make the performance of many alternatives context dependent.

In a gray-tone image rendition of canopy height data obtained from high-intensity small-footprint LIDAR (≥ 1 first return pulses per m^2) the aforementioned 'height' clusters are usually visible; a manual delineation would seem straightforward. However interpersonal differences in interpretation would raise questions about the quality of a manual delineation.

In this study we propose a semi-automatic delineation procedure to capture within-stand areas (clusters) with height characteristics distinctly different from those of the remaining stand. The approach is as follows: An interpreter of a LIDAR gray-tone rendition of a stand first decides on the presence or absence of height clusters in the stand. In the affirmative case the interpreter provides an initial estimate of the number of spatially disjoint clusters in the stand. This estimation can be accomplished quickly and with a minimal rule-set. Once the presence and the number of clusters is decided, the canopy height data is transformed to a three-dimensional array of location and height-class of presumed trees. Next a non-parametric maximum likelihood separation of feature and clutter points is done on data from a selected height-class (Allard and Fraley, 1997). We illustrate this method with an example using data from airborne LIDAR data collected in 2001 on Vancouver Island, Canada over mixed stands of Douglas-fir and Hemlock subject to variable retention cutting.

2. METHODS AND ALGORITHMS

2.1 Feature and clutter points

The clusters to be delineated are composed of a subset of the assumed tree heights that meet specified criteria. The spatial locations of these observations forms a spatial point pattern with zero, one, or more apparent clusters. A cluster is simply a compact spatial sub-domain within the stand where the density of observations meeting the set criteria is significantly elevated compared to the density in surrounding areas. A point inside a cluster is called a feature point while all others are regarded as clutter points. The spatial delineation of a height-cluster is therefore accomplished by a classification of the points as either feature or clutter (Hartigan, 1975). Once classified the spatial clusters are formed by the union of feature points and their associated areas. The definition of the area associated with a point is critical for the actual delineation of clusters; here it is the set of locations closer to a feature point than to any clutter point (Allard and Fraley, 1997).

2.2 A profile likelihood for separating feature and clutter points

Consider a bounded domain *K* with area |*K*| in $\Re^2$ on which we observe a random sample $\{\mathbf{x}_i\}$ of size *n* from a mixture of two uniform random variables: features (U_A) with support $A \subset K$ with probability *p* and clutter (U_K) with support *K* and complementary probability 1-*p*. Both the support *A* and the mixture parameter *p* are unknown and are to be estimated simultaneously from the observed point pattern. The density function associated with a point $\mathbf{x} \in K$ is

$$f(\mathbf{x}) = \frac{p}{|A|} \times \delta_A(\mathbf{x}) + \frac{1-p}{|K|} \quad (1)$$

where $\delta_A(\mathbf{x})$ denotes the feature indicator function for which we have $\delta_A(\mathbf{x}) = 1$ if x is a point in *A* and 0 if it is not

We cannot solve simultaneously for maximum likelihood estimators of *A* and *p*. Instead, if *A* was fixed, the maximum likelihood estimator of *p* is $\hat{p} = \left(\#(A) - |A| \times n\right) / \left(n - |A| n\right)$ and the partial likelihood is obtained accordingly.

A maximum profile likelihood estimator of *A* can only be obtained if we impose constraints on *A* with regard to its shape and number of clusters, otherwise the likelihood would be unbounded. We constrain $\hat{A}$ to be defined by a subset of the *n* Voronoï cells defined by the data in the bounded domain *K* (Okabe et al, 2000). A subset of spatially connected Voronoï cells forms a cluster.

For a fixed number of Voronoï cells, say *m*, the sub-region maximizing (3) is the union of the *m* smallest cells. $\hat{A}$ is therefore determined by the set of the *m* smallest Voronoï cells maximizing (3). If $\hat{A}$ contains more than the prescribed number of clusters (say *C*) only the *C* largest clusters are retained in $\hat{A}$, the others are disregarded.

2.3 Regularizing the shape of clusters by morphological closure

The natural boundary of a cluster may not be well approximated by the procedure outlined above. The purely mathematical definition of a Voronoï cell enhances the risk of a highly irregular ('unnatural') outline of connected subsets; some may even contain 'holes'. A regularized outline and plugging of holes may be desired in practical applications. The closing operation on graphs in mathematical morphology can be used to 'smooth' the outline of a cluster to a desired degree of regularity (Pratt, 2001).

2.4 Clustering of height quartile classes

The spatial locations of presumed trees in the first quartile of an assumed tree height distribution recovered from LIDAR canopy height data in a stand dominated by Douglas-fir on Vancouver Island (British Columbia, Canada) are shown in Figure 1. The LIDAR data were collected by Mosaic Mapping Systems Inc. using a Riegl Q140i 60 laser with a wavelength of 0.9 µm. The target flying altitude was 220 m AGL resulting in a swath width of approximately 220 m. The scan angle was set with an allowance of up to +/- 30 degrees. A flight overlap of 40 percent was the result of a flight line spacing of 120 m. Using a scan rate of 34 Hz a point density of approximately 0.4 m x 1.0 m resulted. The beam divergence was programmed for 3.0 mrads (resulting in a spot size of 90 cm at 300 m flying altitude). The collection mode of the data was for first and last return and intensity. The conversion from canopy height data to assumed height of individual trees was achieved in two steps. In step one the distribution of canopy height differences between the maximum value and all other values within a window of size $\lceil 0.24 \times meCH^{1.68} \rceil$ was computed ($meCH$ is the median stand LIDAR canopy height, (Magnussen and Boudewyn, 1998). In step two the stand is gridded with a grid-size equal to the smallest integer larger than the mode of this distribution and the maximum LIDAR canopy height in each grid cell is retained as the height of a presumed tree. The spatial location of this maximum in the grid cell is retained as the spatial location of this tree.

A visual inspection of the point pattern indicates a non-random spatial distribution. We assume that trees in the first height quartile are clustered into three disjoint clusters. The maximum profile solution in Figure 1 and the smoothed solution in Figure 2 share many common features with the latter distinctly more attractive.

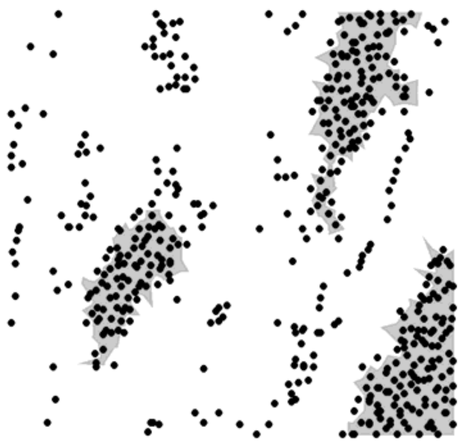

Figure 1. Maximum profile likelihood estimator of three clusters (shaded area).

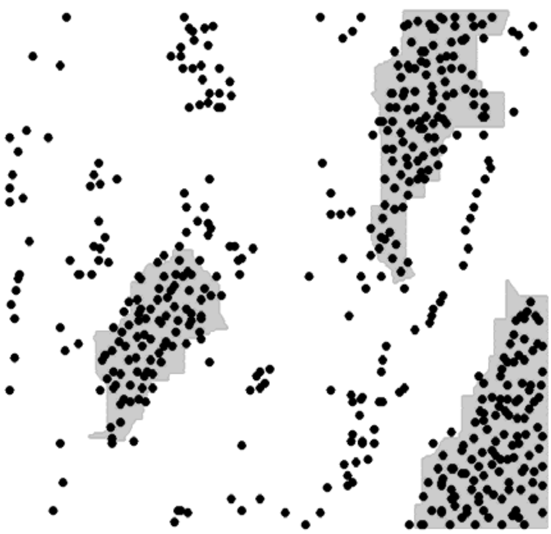

Figure 2. Smoothed maximum profile likelihood estimator of three clusters (shaded area).

Associated cluster point densities were, in both cases, about twice as high as the density of clutter points and no cluster was less than 0.2 ha.

3. DISCUSSION AND CONCLUSIONS

The rapid increase in the average density of first-return LIDAR canopy heights captured over forest stands opens new opportunities for the forest inventory analyst beyond recovery of tree heights and associated tree attributes like volume and biomass. Tools and methods available for image analysis apply increasingly also to LIDAR data.

Clustering of point patterns can be achieved in a variety of ways, from segmentation of a grayscale 'images' of canopy heights (Hill, 1999; Lee, 2000) to a classification based on nearest neighbor distances (Byers and Raftery, 1998). The nonparametric likelihood approach rests on a minimum of assumptions and is computationally straightforward. Allard and Fraley (Allard and Fraley, 1997) pioneered the method also and variations in the way cells were added or removed from the current set of Voronoï cells under consideration for clustering. They found the solutions obtained after closure with Voronoï cells as structuring elements was the best overall choice.

We conclude that the nonparametric profile likelihood approach to delineation of clusters in spatial point processes is suitable for special applications in forest inventory and provides an additional, and unique, option for the analysis of LIDAR data.

REFERENCES

Allard, D., and Fraley, C. 1997. Nonparametric maximum likelihood estimation of features in spatial point processes using Voronoi tesselation. *Journal of the American Statistical Association,* 92, pp. 1485-1493.

Byers, S.D., and Raftery, A.E. 1998. Nearest-neighbor clutter removal for estimating features in spatial point processes. *Journal of the American Statistical Association,* 93(442), pp. 577-584.

Cho, V.H.K. 2004. Investigations for segment-based classification of satellite images for the purpose of forest mapping. *Allgemeine Forst- und Jagdzeitung,* 175(4-5), pp. 94-100.

Diggle, P.J. (1983). *Statistical Analysis of Spatial Point Patterns*. Academic Press, New York, pp. 148.

Hartigan, J.A. (1975). *Clustering algorithms* . Wiley, New York, pp. 351.

Hill, R.A. 1999. Image segmentation for humid tropical forest classification in Landsat TM data. *International Journal of Remote Sensing,* 20(5), pp. 1039-1044.

Lawson, A.B., and Denison, D.G.T. (2002). *Spatial cluster modelling*. Chapman & Hall (CRC), Boca Raton, FL., pp. 226.

Lee, T.C.M. 2000. A minimum description length-based image segmentation procedure, and its comparison with a cross-validation-based segmentation procedure. *Journal* *of the American Statistical Association,* 95(449), pp. 259-270.

Magnussen, S., and Boudewyn, P. 1998. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research,* 28, pp. 1016-1031.

Okabe, A., Boots, B., Sugihara, K., and Chiu, S.N. (2000). *Spatial tesselations. Concepts and applications of Voronoi diagrams*. (ed. 2). Wiley, Chicester, pp. 696.

Pratt, W.K. (2001). *Digital Image Processing* . John Wiley & Sons, New York, pp. 758.

Ripley, B.D. 1985. Statistics, images, and pattern recognition. *Canadian Journal of Statistics,* 14(2), pp. 83-111.

Upton, G.J.G., and Fingleton, B. (1985). *Spatial data analysis by example. Volume I. Point patterns and Quantitative data*. Wiley, Chichester, pp. 468.