IDENTIFYING AND QUANTIFYING HETEROGENEOUS BOREAL FOREST STRUCTURES USING LASER SCANNER DATA

Maltamo, M.[a], Packalén[a], P., Yu, X. [b], Eerikäinen, K. [c], Hyyppä, J. [b] and Pitkänen, J. [a]

a) University of Joensuu, Faculty of Forestry, P.O. Box 111, FIN-80101 Joensuu, Finland. Tel.+358 13 2513615. E-mail: matti.maltamo@forest.joensuu.fi, petteri.packalen@forest.joensuu.fi,juho.pitkanen@forest.joensuu.fi
b) Finnish Geodetic Institute, Geodeetinrinne 2, P.O. Box 15, FIN-02431 Masala, Finland, E-mail: yu.xiaowei@fgi.fi, juha.hyyppa@fgi.fi
c) Finnish Forest Research Institute, Joensuu Research Centre, P.O.Box 68, FIN-80101 Joensuu, Finland. E-mail: kalle.eerikainen@metla.fi

KEY WORDS: remote sensing, forestry, interpretation, inventory, recognition, lidar, structure

ABSTRACT

Structural characteristics of forest stands, e.g. in relation to carbon content and biodiversity, are of special interest. It has been stated in numerous publications that discrete return laser scanner data produce accurate information on a tree canopy since the quantiles of the height distribution of laser scanner data are related to the vertical structure of the tree canopy. Since some of the laser pulses will penetrate under the dominant tree layer, it is also possible to analyse multi-layered stands. In this study the existence and number of suppressed trees was examined. This was carried out by analysing the height distributions of reflected laser pulses. In this laser data (Toposys Falcon, survey May 2003 in Kalkkinen, flight altitude 400 m AGL) the average number of laser pulse hits per 1 m^2 was 12. The reference data consisted of 28 accurately measured field sample plots. These plots include highly heterogeneous boreal forest structures. The existence of lower canopy layers, i.e. suppressed trees, was analysed visually by viewing the plotwise 3D-images of laser scanner based canopy height point data and examining distributions of canopy densities which were computed as the proportions of laser hits above different height quantiles. Furthermore, a developed histogram thresholding method (HistMod) was applied to the height distribution of laser hits in order to separate different tree storeys. Finally, the number and mean height of suppressed trees were predicted with estimated regression models. The results showed that multi-layered stand structures can be recognised and quantified using quantiles of laser scanner height distribution data. However, the accuracy of the results is dependent on the density of the dominant tree layer.

1. Introduction

In Finland, managed boreal forests are structurally rather homogeneous due to silvicultural operations and low tree species diversity. The planting of trees has favoured coniferous tree species and silvicultural operations have removed suppressed trees and tree groups. These operations have usually caused a uniform stand structure in boreal forests. On the other hand, unmanaged boreal forests are often more heterogeneous including multi-layered tree canopies and uneven-aged stand structures (e.g. Esseen et al. 1997). In Finland, such forests are nowadays rare. However, their importance for the biodiversity of forests and the protection of endangered species is undisputable. Some natural structural characteristics may still be found from old managed spruce forests whereas pine dominated forests are usually rather homogeneous.

It is almost impossible to examine vertical stand structure using optical remote sensing imageries. On the other hand, it has been stated in numerous publications that discrete return laser scanner data produce accurate information on a tree canopy since the quantiles of the height distribution of laser scanner data are related to the vertical structure of the tree canopy (e.g. Magnusseen & Boudewyn, Naesset 2002). Naesset (2004) has also visualised the relationship between stand structure and laser point height distribution. Until now, most of the small-footprint lidar studies have focused on analysing the structures of single-layered forests (e.g. Zimble et al. 2003) although lidar based fire behaviour models may include information also from understorey (Riano et al. 2003). Furthermore, the waveform-recording laser altimetry (SLICER) has been applied also to the characterisation of multi-layered forests (e.g. Harding et al. 2001).

The aim of this study is to examine heterogeneous boreal forest structures in a boreal nature reserve using lidar canopy height point data. Since some of the laser pulses will penetrate under the dominant tree layer, they are used to characterise the existence and number of suppressed trees.

2. Material and methods

2.1 Field data

The test site was placed in a state owned forest area of approximately 50 hectares located in Kalkkinen, southern Finland, 130 km north of Helsinki. The tree stock of the area is naturally regenerated and there have been no silvicultural operations in most parts of the area during the last decades. Therefore, the tree stock can be considered to be in a semi-natural state. During the summer of 2001 28 systematically located rectangular sample plots were established on the test site. The basic size of a sample plot was 30 m by 30 m, but to get around 100 trees per plot, plot sizes of 25 m by 25 m and 30 m by 40 m were also used. Most of the sample plots included dense understorey and were dominated by Norway spruce, although, each sample plot included more than one tree species. All trees having a diameter at breast height (DBH) of more than 5 cm were mapped and tree species, DBH, tree height, and height to the living crown were registered. All individual trees were classified as belonging to either the dominant or the suppressed tree layer. On the average there were 414 dominant and 435 suppressed trees per hectare in the data. Corresponding figures for Lorey's mean height were 26 and 12 m, respectively. In addition, it was also determined on the plot level whether there are one or more tree storeys. Altogether 18 plots were classified as multi-layered.

Subsequently, the Global Positioning System (GPS) was used to determine the position of the four corners of each of the 28 sample plots. The coordinates of the corner points were measured with a Leica SR530 RTK GPS system with an RTK mode or as a static GPS measurement. RTK is based on the concept of relative GPS where two receivers, one at a reference point with known coordinates and one at a new point, simultaneously measure signals from the same satellites.

2.2 Laser scanning

Using a Toposys Falcon scanner, georeferenced lidar point cloud data were collected from Kalkkinen on 15 May 2003. Three DGPS receivers were employed to record the carrying platform position: one on the aircraft, and two on the ground (the first as the base station, the second for backup). The laser scanner survey provided a point cloud, in which the x, y and z coordinates of the points are known. The test site was measured from an altitude of 400 m (above ground level, a.g.l.) resulting in a nominal sampling density of about 10 measurements per m^2. Due to the relatively low survey altitude applied, the swath width was approximately 100 m. First pulse data were applied since previous studies have shown that first pulse data can be successfully used even for digital terrain model (DTM) generation in a boreal forest zone. Laser point clouds were first classified by TerraScan software (see www.terrasolid.fi) to separate the ground points from other points. Then a raster DTM grid with a 50 cm pixel size was created from classified ground points by taking the mean value of the ground points within the grid. Missing points in the DTM were afterwards interpolated using Delaunay triangulation and the bilinear interpolation method. Laser canopy heights were calculated as the difference between z values of laser hits and estimated ground elevation values at the corresponding location. Ground elevation values of laser hits were interpolated from created. Points which canopy height value was over 0.5 meters were expected to be vegetation hits.

2.3 Methods

The existence of lower canopy layers, i.e. suppressed trees, was first analysed visually by viewing the plotwise 3D-images of laser scanner canopy height point data and comparing them to the field data. Different height metrics were calculated from the laser canopy height point data by sample plots. Percentiles for the canopy height for 0, 5, 10, 20, …, 90, 95 and 100 % were computed. Cumulative proportional canopy densities were calculated above different height quantiles at the corresponding points. Finally, proportion of vegetation hits were computed. Distribution of canopy density were used to analyse the stand structure and all calculated metrics were used in the prediction model of number and mean height of suppressed trees.

An algorithm referred here as HistMod was used to analyze whether the height distribution of laser hits is multimodal or not. If the height distribution of laser hits is considered to be multimodal, underlying canopy structure is considered to be multi-layered. The fundamental part of the HistMod algorithm is to use a method presented by Lloyd (1982) as a first iteration to define a threshold of dominant tree layer and under storey trees. Lloyd's algorithm is an *N*-level vector quantizer for *k*-dimensional blocks of samples. In vector quantization the goal is to minimize the squared quantization error by grouping similar data vectors together. In the case of height distribution of laser hits (i.e. frequency histogram), *k* is 1, and because the purpose is only to separate dominant tree layer and under storey trees, *N* is 2. Thus, all data vectors are assigned either to class 'dominant tree layer' or 'understorey tree layer', separated by threshold value. When *N*=2 and *k*=1 Lloyd's threshold (*T*) is calculated as follows:

Initialize $T = \bar{x}$
REPEAT
    Threshold the histogram using T
    Calculate class averages $\bar{x}_1 = \frac{1}{n_1}\sum_{x_i \le T} x_i$ and $\bar{x}_2 = \frac{1}{n_2}\sum_{x_i > T} x_i$
    Recalculate threshold $T = \frac{\bar{x}_1 + \bar{x}_2}{2}$
UNTIL T does not change

The algorithm starts by taking the average values ($\bar{x}$) as a tentative threshold (*T*). The histogram is divided into two parts by using the threshold *T* and the class averages $\bar{x}_1$ and $\bar{x}_2$ are calculated (n_1 and n_2 are the number of elements). New threshold value is taken as the average of $\bar{x}_1$ and $\bar{x}_2$ and the process is repeated until the threshold value *T* does not change. The final threshold value divides the histogram into two classes.

The HistMod algorithm uses the following procedure to define the multimodality of height distribution of laser hits (See also Figure 3.):

1. Exclude vegetation hits and create a height distribution of laser hits within a sample plot to be examined. Hereinafter this is referred as the histogram. In this study the used bin width was 1 cm.
2. Low-pass filter (moving average) the histogram with a kernel width K_{width}. In this study the histogram was filtered twice to get rid of high frequency fluctuation.
3. Define Lloyd's threshold as explained above.
4. Move from the Lloyd's threshold descent until the slope is zero within a kernel SK_{width}. This point is denoted as *UpperBound*.
5. Search the highest point from the left side of the *UpperBound*. This point is denoted as *MaxFreq*.
6. Search the lowest point between the *MaxFreq* and *UpperBound*. This point is denoted as *MinFreq*.
7. Calculate the difference of frequencies (*DiffFreq*) between the *MaxFreq* and *MinFreq*.
8. If *DiffFreq* is greater than a given difference limit *DiffLim*, histogram is considered to be multimodal, i.e. a plot is considered to have a multi-layered canopy structure.

3. Results

An example of the 3D-laser canopy height point cloud with the tips of the field measured trees in one sample plot is presented in Fig. 1. The multi-layered stand structure can be clearly seen from the figure.

The stand structure can also be analysed in relation to vertical cumulative distributions of canopy densities (Fig 2.). According to Korpela (2004) trees whose height is more than 60 % of stand maximum height can be seen and recognised from optical high-resolution imageries in Finnish conditions. If this value is correspondingly used as a threshold to analyse the existence of lower canopy layers the differences between different plots, e.g. single-layered plot A and multi-layered plot B, can be observed. However, the results are also dependent on the density of the dominant tree layer. In a forest with very dense dominant layer (plot C), the proportion of laser hits in lower layers is minor, although there exists a dense understorey.

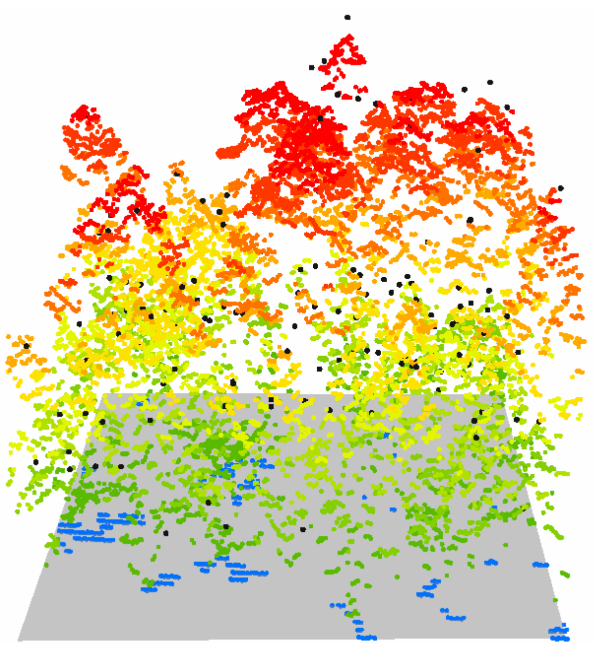

Figure 1. An example 3D-laser canopy height point cloud in multi-layered sample plot. Field measured tree tips (black dots) are also shown.

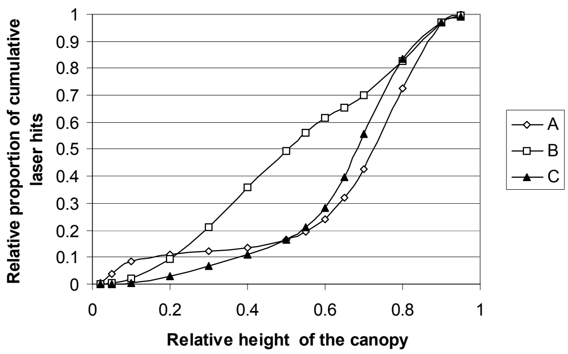

Figure 2. Examples of vertical cumulative distributions of laser hits. Plot A = single-layered stand, B = multi-layered stand, C = multi-layered stand with very dense dominant tree layer

The HistMod algorithm was applied to all 28 sample plots of which 24 were classified correctly to have a single- or multi-layered canopy structure (Fig. 3). Four sample plots were misclassified to have a single canopy layer although these plots are multi-layered. The algorithm was executed numerous times with different parameters to find an optimal parameter combination. The best parameters proved to be K_{width} = 181 cm, SK_{width} = 10 cm and *DiffLim* = 1.0 frequency. Note that a narrow bin width results in a low optimal value of *DiffLim*. Parameters K_{width} and *DiffLim* are highly dependent, whereas the algorithm seems to be rather insensitive to the parameter SK_{width}.

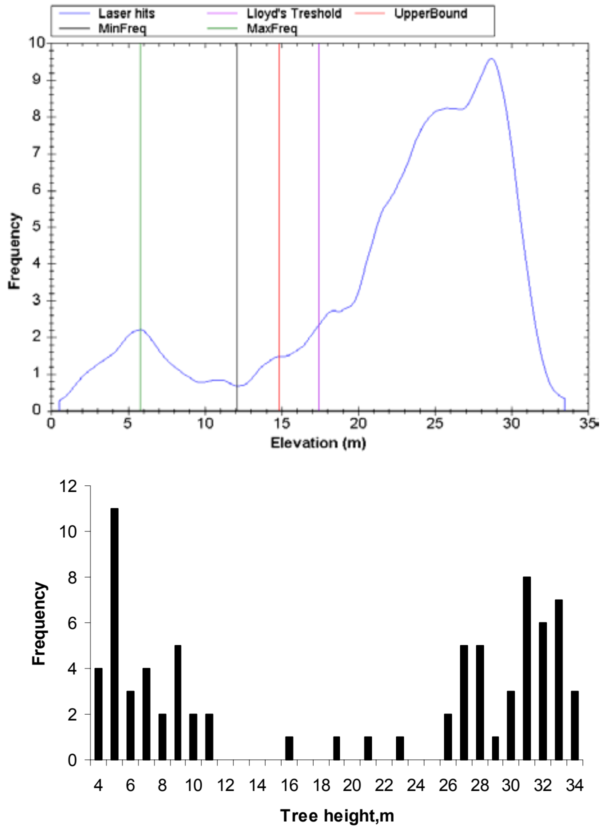

Figure 3. Upper: Example of the HistMod algorithm procedure used to define the laser hit distribution in multi-layered stand. Lower: Corresponding field measured tree height distribution.

The constructed model for the logarithmic number of suppressed trees is presented in Table 1. The reliability figures of the model are as follows: R^2 of the model is 0.87 and standard error 0.312. It is worth noting, that height metrics of lower part of the canopy are used as independent model variables.

Predictor	Coefficient	Standard error	t-value
Constant	-4.268	2.538	-1.682
proportion of vegetation hits	10.605	0.868	12.212
canopy height percentile 20 %	-0.153	0.039	-3.944
lidar maximum height	-0.171	0.049	-3.527
ln(canopy height percentile 60 %)	2.650	1.328	2. 000

Table 1. Regression model for the logarithmic number of suppressed trees

The only independent model variable used in the prediction model for the logarithmic Lorey's mean height of suppressed trees was logarithmic maximum height of suppressed trees (coefficient 0.800, standard error 0.089 and t-value 8.97). Corresponding values for the intercept of the model are

0.209, 0.250 and 0.84, respectively. The reliability figures of the model are as follows: R^2 of the model is 0.76 and standard error 0.13.

4. Discussion

This study examined the use of canopy height laser point data to characterise vertical stand structure of heterogeneous boreal forests. It was found out that characteristics of canopy height laser point data, especially the shape of the height distribution, can be utilised to recognise multi-layered stand structures. Visual check of 3D laser images, analysis of cumulative vertical distributions of canopy densities and histogram thresholding method were used to recognise features of multi-layered forests.

The HistMod algorithm was used to test whether the canopy height distribution of laser hits is multimodal or not. If the height distribution of laser hits was recognized to be multimodal then the canopy structure was also considered to be multi-layered. The classification succeeded in 24 sample plots and failed in 4 sample plots. In all sample plots where the classification failed the canopy height distribution of laser hits was not evidently multimodal although the underlying height distribution of trees was multimodal. Thus, HistMod classifies plots correctly if the multi-modality is shown in the canopy height distribution of laser hits. Unfortunately this is not always the case since the multi-modality of laser canopy height point distribution is dependent on the density of dominant tree layer.

Suppressed trees were quantified using regression models for the number and mean height of these trees. These characteristics can be further used to predict the volume of small trees. It was found out that the model for the number of trees is relatively reliable and usable containing characteristics of lower part of laser canopy height distribution. On the other hand, modelling the mean height was found to be problematic; the only independent model variable is maximum height of suppressed trees. An estimate for that characteristic could be obtained from the HistMod algorithm. The applicability of such approach needs, however, further research. In addition, it is not always necessary to predict volumes of suppressed trees. Sometimes it may be more useful only to recognise the occurrence of these trees.

The density of laser hits per m^2 was high enough in this data to allow recognition of individual trees also (see e.g. Maltamo et al. 2004). However, only trees of dominant tree layer were found and in multi-layered stands it results in an inadequate description of the stand structure. For such situations Maltamo et al. (2004) proposed the combination of pattern recognition of big trees and theoretical distribution function for small trees. Height distribution of small trees was predicted using information obtained from dominant trees only. The approach of this study is, however, more advanced since it utilises laser hits reflected from lower tree layers and uses this information in the prediction of suppressed trees.

For large area practical inventories the point density of the data of this study is probably too expensive to measure. In one plot, which area was about 900 m^2, there were about 10 000 laser hits. In large area inventories the basic unit is a stand which mean size in Finnish conditions is about 2 hectares. When low pulse densities of about 1 hit per m^2 are used, the number of laser hits obtained at stand level may still be high enough for the recognition of the occurrence of lower tree layers (see also Naesset 2004).

The approach of this study can be recommended for mapping structurally diverse forest areas. In addition to that spatially more detailed description of tree stock could be utilised in laser scanning based biomass and carbon content studies. Finally, also a need for silvicultural operations (e.g. removal of shelterwoods) and their timing could be examined using vertical laser canopy height point distributions.

Acknowledgements

This study is a part of the project "The usability of single tree laser scanning in forest planning (decision number 78238)" funded by Academy of Finland.

References

Esseen, P.-A., Ehnström, B., Ericson, L. & Sjöberg, K. 1997. Boreal forests. *Ecological Bulletins*, 46. pp. 16-47.

Harding, D.J., Lefsky, M.A., Parker, G.G. & Blair, J.B.2001. Laser altimetry canopy height profiles methods and validation for closed-canopy, broadleaf forests. *Remote Sensing of Environment,* 76. pp. 283-297.

Korpela, I. 2004. Individual tree measurements by means of digital aerial photogrammetry. *Silva Fennica Monographs,* 3. 93 p.

Lloyd, S.P., 1982. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), pp. 129-137.

Magnussen, S., & 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.

Maltamo, M., Eerikäinen, K., Pitkänen J., Hyyppä, J. and Vehmas, M. 2004. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. *Remote Sensing of Environment,* 90. pp. 319-330.

Næsset, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment,* 80. pp. 88-99.

Næsset, E. 2004. Practical large-scale forest stand inventory using a small airborne scanning laser. *Scandinavian Journal of Forest Research,* 19. pp. 164-179

Riano, D., Meier, E., Allgöwer, B., Chuvieco, E. and Ustin, S.L. 2003. Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. *Remote Sensing of Environment*, 86. pp. 177-186.

Zimble, D.A., Evans, D.L., Carlson, G.C., Parker, R.C., Grado, S.C. & Gerard, P.D. 2003. Characterizing vertical forest structure using small-footprint airborne LIDAR. *Remote Sensing of Environment,* 87. pp. 171-182.