AUTOMATIC DELINEATION OF FOREST STANDS FROM LIDAR DATA

V. J. Leppänen [a, *], T. Tokola [a], M. Maltamo [a], L. Mehtätalo [a], T. Pusa [b], J. Mustonen [c]

[a] University of Joensuu, Faculty of Forest Sciences PL111, 80101 Joensuu, FINLAND - wleppanen@gmail.com
[b] Arbonaut Oy, Koskikatu 5 C, 80100 Joensuu, FINLAND – tuomo.pusa@arbonaut.com
[c] University of Helsinki, Faculty of Agriculture and Forestry,PO.Box 27, 00014 University of Helsinki, FINLAND – jukka.mustonen@helsinki.fi

KEY WORDS: LiDAR segmentation, automatic stand delineation, stand definition, stand quality, forest stand delineation, automatic feature extraction, region growing

ABSTRACT:

Stands are the basic unit of forest management and information. The main disadvantage of stands is the labour-intensive updating work needed and the high variance in the quality of the results due to the amount of subjective judgment and manual work in creating the stands. In this study, a new approach for automatic stand delineation from a composite of LiDAR data –derived raster layers and colour-infrared aerial imagery –derived layer is introduced. The **segmentation utilizes a new, iterative region-growing -based approach that forces the stands to be homogenous in timber type**.

The stand delineation quality was compared based on their ability to separate timber characteristics. The developed method was compared to two other stand delineation methods:

- Automatically interpreted, with a segmentation algorithm in eCognition Pro 4.0 using LiDAR canopy height model.
- Human-interpreted on aerial imagery (The traditional way in Scandinavian forestry).

The testing was done on a 67-hectare forestland area in Juuka, Finland. 683 sample plots were laid on the property for control. This research shows it is possible to produce stand delineation automatically, utilizing LiDAR data, if timber characteristics are the only stand boundary criteria considered.

1. INTRODUCTION

Forest stands are used as timber inventory units, forestry data containers and operation units in timberland management. Stands can be made either on operational or on biological basis. Biological stands tend to be smaller and more detailed than stands made on operational basis. Typical stand criteria have been timber size, density and species, as well as site type.

Traditionally, the stands have been produced using aerial imagery in various formats, like printed film, stereo pair or an orthoimage. Stand delineation has been done mainly manually, even though some automatic approaches using for example aerial orthoimagery or LiDAR data have been used. The main disadvantage of using the stands is the labour-intensive updating work needed and the high variance in the quality of the results due to the amount of subjective judgment and manual work in stand production (Haara & Haarala 2002).

Automatic stand delineation algorithms using color-infrared orthoimagery (CIR) have been developed in several commercial and research projects (Pekkarinen 2001, Sell 2002, Leckie et al. 2003). CIR, a commonly-used material for manual and automatic stand delineation, has signal about tree species variation, especially differentiating hardwood and conifers, but lacks information about timber height.

LiDAR data has been used in many forestry applications. Hyyppä et al. (2001) presented the idea to predict timber characteristics using individual trees segmented from high-resolution LiDAR. Maltamo et al. (2004) further developed the methodology for receiving more accurate diameter distributions. Naesset (2002) developed a process for predicting total timber volume, basal area, height, stem count and diameter at breast height on a grid, laid over the forest land, using a set of LiDAR-derived independent variables and a ground sample.

LiDAR data provides a good basic material also for stand delineation, but it is lacking a clear signal of species information. Segmentation method using LiDAR-derived vegetation height has been introduced by Mustonen (2007). LiDAR has been used for delineation of man-made structures (Wang & Tseng 2004). Diedershagen et al. (2004) used commercial FOGIS -software and high-resolution LiDAR combined with multispectral data to produce forest stands. The approach was to input high-resolution digitized vegetation height model and high resolution data into the segmentation algorithm.

Estimating the quality of stand delineation has been difficult, since the delineation is a result of judgments on a series of issues. Especially difficult is the quality comparison on operationally based stands. Mustonen (2007) used analysis of variance on the stand quality comparisons using, essentially, biologically based stand delineation.

In this study, three approaches for stand delineation LiDAR data

* Corresponding author. This is useful to know for communication with the appropriate person in cases with more than one author.

and CIR are compared. First approach uses 1 meter resolution tree crown height model (CHM) introduced to commercial multi-resolution segmentation tool (eCognition). Second approach, LIRGA algorithm, utilizes LiDAR-derived layers with high correlation to vegetation height and density, composited to CIR-derived layer with high correlation to hardwood content of each analysis cell. The analysis resolution used in segmentation in this approach is 4 meters. A specialized segmentation technology, designed to maintain segment homogeneity in low-noise input raster is used. The two automatic delineation methods were compared with third method, traditional forestry specialist-made operational stand delineation, commonly used in Scandinavian forestry.

2. DATA AND PREPROCESSING

The research area is a 67 ha commercial forest property owned by United Paper Mills and located in Juuka, Eastern Finland. It has been managed in a manner typical of Scandinavian conditions. LiDAR data were acquired on July 13, 2005 with an Optech ALTM 3100C sensor. The nominal average point density was 0.6 pulses/ m2, varying in the range 0.5-1 pulses/m2. The flight altitude was 2000m above ground level and the field of view 30 degrees, with a 60 cm beam footprint. Four returns were recorded by the sensor, and the first and last pulses were attributed. The returns were classified into two classes: “ground” and “default”. A 2.5 m DTM was created using the mean of the ground returns as the z-value and bilinear interpolation for the cells with no ground returns. A vegetation height model was made from the LiDAR return point cloud by replacing the z-values with the difference between the point and the DTM altitudes.

For validation, a systematic grid sample of 729 plot centers with 30-metre spacing was laid over the area. Some of the plots landed on non-forested land, leaving 683 plots to be measured, which was done in July 2006. The plot density was 9.6 plots/ha. The site class was estimated and timber characteristics were measured by species and by canopy layer if several canopy layers existed within one species. A relascope with a multiplier of 2 was used for measuring basal area and selecting the sample trees for dbh and height measurements on timbered plots. The dbh of the basal area median tree was measured with calipers and the height of the same tree using a Vertex height measurement device. In seedling areas a circular 50-m2 fixed area plot was used. A Pathfinder ProXRS GPS device with real-time differential corrections was used for measuring the locations of the plot centers. The timber characteristics of the plots were calculated by generating a beta-function to estimate the stand diameter distribution for each species and canopy layer. The volume characteristics of the trees on each plot (Pukkala 2004) were calculated using the height model of Siipilehto (1999) and standard volume taper curves.

Table 1. contains definitions of the plot data variables, their means, population variances and minimum and maximum values.

Table 1. The stand characteristics produced for the control plots. Units are in m^3/ha.

Variable	Unit	Population mean	Population variance	Population Minimum	Population Maximum
Total					
Basal area	m^2/ha	18	145	0	58
Diamater at breast height	cm	19.0	74	0	34.5
Height	m	15.6	45	0	23.9
Volume	m^3/ha	132	9373	0	566
Stemcount	count	426	510546	0	12600
Scots Pine					
Basal area	m^2/ha	12	105	0	44
Diamater at breast height	cm	21.3	114	0	46.5
Height	m	16.9	59	0	24.9
Volume	m^3/ha	95	8105	0	364
Stemcount	count	321	1586426	0	10000
Norway Spruce					
Basal area	m^2/ha	4	65	0	56
Diamater at breast height	cm	15.1	50	0	33.0
Height	m	13.0	35	0	22.1
Volume	m^3/ha	26	3336	0	550
Stemcount	count	42	154103	0	9000
Hardwoods					
Basal area	m^2/ha	2	13	0	24
Diamater at breast height	cm	12.2	37	0	26.9
Height	m	12.7	34	0	21.0
Volume	m^3/ha	11	580	0	185
Stemcount	count	62	98684	0	3000

3. METHODS

3.1 Segmentation methods

3.1.1 Hierarchical segmentation on crown height model

CHM segmentation is based on eCognition Pro’s hierarchical network of image objects. Each image object is connected both to upper and lower level image objects. This allows the consideration of the final segments onto the finest segmentation level, like recognized individual trees in the area. For the best operation of the hierarchical network, it is important to feed the algorithm with a high resolution CHM data, which allows the analysis of finer crown structure.

CHM segmentation was applied on a 1 meter resolution CHM raster image. The first step in the creation of CHM image was creation of an empty 1 meter resolution raster set over the study area. The raster was initialized with value -1. Secondly, the algorithm searched the highest laser point within the radius of 2 m (Euclidean distance) from the center of each pixel. Finally, pixels remaining empty (pixel value -1) were filled by calculating the average of the laser points of the neighbouring pixels. Filling was done recursively so that the new value was given only to those pixels that had at least four non-empty neighbours. Elevation values in the CHM are vegetation hits, so the values close to zero are returns from ground, while the highest values are returns from vegetation. Water bodies were masked out of the CHM. Figure 1 presents the CHM input to the algorithm.

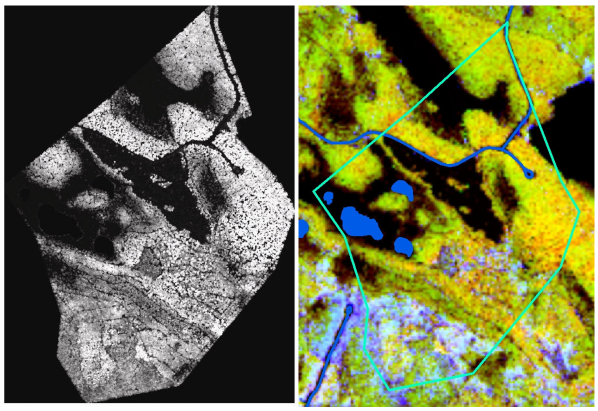

Figure 1. CHM input for hierarchical segmentation approach on the left. Height–density-hardwood percentage –input for the LIRGA algorithm on the right.

3.1.2 Limited iterative region growing on composite of LiDAR and CIR –derived feature rasters

The second segmentation approach used is called a "limited iterative region growing algorithm" (LIRGA) due to the idea of limiting the region-growing unless pixels similar enough to the seed environment are found. Unless most of the region growing algorithms, this one is designed to be conservative in the seed finding phase. First, the algorithm finds some promising seed points, segments out the homogenous areas around those seeds by region growing algorithm that allows growth only if the values of new prospective pixels to join the segment are close to the means of the seed vicinity. After finding first phase segments, the algorithm iterates by adding new seeds only on the unsegmented areas by gradually decreasing the seed-finding criteria and re-runs the region growing, starting from combined seeds from all previous iterations.

A composite of three bands was used for LIRGA input. The bands were selected to provide best possible correlation with the stand-defining criteria while being simple to compute from the input data and easy to adjust in new conditions. Band1 was a 85^{th} height percentile of the LiDAR points, representing a height value with 85% of the LiDAR returns below the value, was used as the value of this band. The height analysis was done on a grid of 8m pixel size to yield a good spatial accuracy while it filtering individual tree-sized features and small canopy gaps out. For better edge accuracy, the resulting raster was interpolated to 4-meter raster using bilinear interpolation function. Band 2 was designed to maximally correlate with the vegetation density, by taking the percentage of LiDAR values returning from vegetation in each cell of an 8-meter grid. For better edge accuracy, the resulting raster was interpolated to 4-meter raster using bilinear interpolation function. Band 3 was targeting to maximally correlate with the percentage of hardwood by volume in the pixel. First, each pixel on 0,5m resolution CIR was classified to hardwood, conifer or non-vegetation using thresholding of band differences. Then, a 4-meter grid was laid over the vegetation class raster, and the percentage of hardwood pixels in each cell was attributed as a pixel value. Figure 1. Presents the composite input to LIRGA algorithm.

The region-growing algorithm above does not necessarily guarantee all neighboring segments would differ from each other in terms of the input band values or variances. A small threshold in the bands may have stopped the segment from growing, and a new seed has been placed later on. However, it is appropriate to merge two adjacent stands if the band mean values and standard deviations are very close to each other. In merging analysis, mean and standard deviation of each band and the sum band were calculated. If all band means and standard deviations were closer than band-unique merging threshold, the merging was performed.

3.1.3 Manual expert-made segmentation on CIR orthoimagery

The third stand delineation strategy was an expert-made operative delineation on CIR imagery. The stand criteria were not exactly the same as in the automatic delineation, since foresters take into account some operative constraints like minimum operative stand size. However, this delineation provides a good benchmark on applicability of the automatic approaches in operative situations.

3.2 Stand map analysis

Analysis of variance was used to test the ability of the delineations to separate different stand types. First, each plot was assigned to a stand on each of the delineations. The means of each stand on the timber volume were calculated as the means of the plots in that stand.

Adjusted R^2's were used to test the ability of stand delineation to separate different timber types.

Additionally, topological line analysis was done for each of the stand lines. Median of timber height and diameter at breast height as well as mean basal area and hardwood percentage out of total volume were considered in this analysis. The means and medians of the neighbouring stands were compared in terms of each of the criteria and if the comparison yielded difference larger than a given threshold, the stand boundary was deemed to be useful in terms of the criteria. Line strength was defined as the number of terms the line was valid. If the line got strength of 0, it was deemed to be unnecessary.

The means of line strengths, weighted by the line lengths, were calculated for each of the delineation for comparison of overall necessities. This comparison is a contradictory force to the adjusted R^2 analysis that, despite of using degrees of freedom, may have tendency to favour a detailed delineation.

4. RESULTS

Three segmentation approaches were compared (Table 2). The hierarchical segmentation approach (CHM) is mainly focused on raw biomass only. The segmentation approach utilizing a composite of LiDAR and CIR –derived features LIRGA improves species-specific results.

The result of applying the CHM delineation is presented in Figure 5. Useful boundaries are colored red, while unnecessary ones are colored green. The mean stand size in the resulting stand map was 0.8 ha.

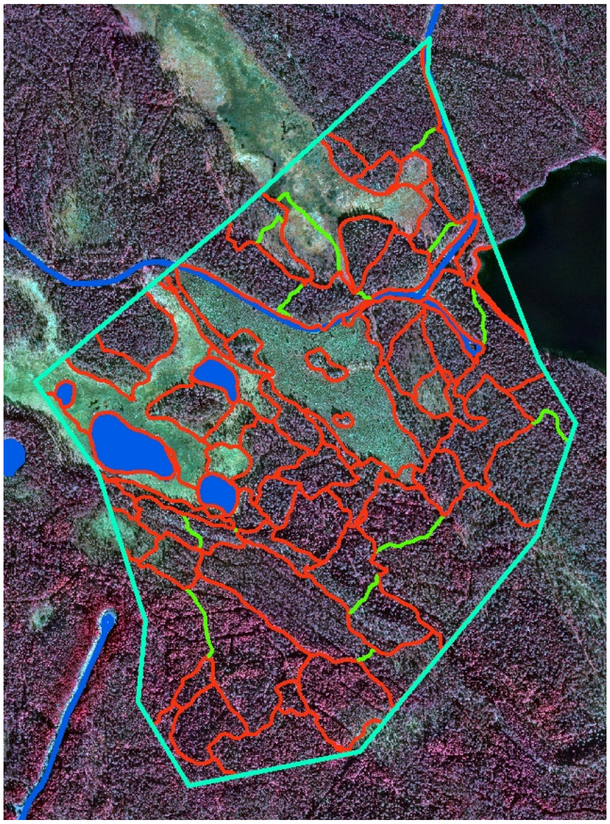

Figure 1. Result of hierarchical segmentation approach based on canopy height model only. Red lines are classified to be strong, while green lines are weakly separating stands.

The result of applying the LIRGA, stand merging and cleaning on H_D_HW raster is presented in Figure 5. The mean stand size in the resulting stand map was 0.7ha.

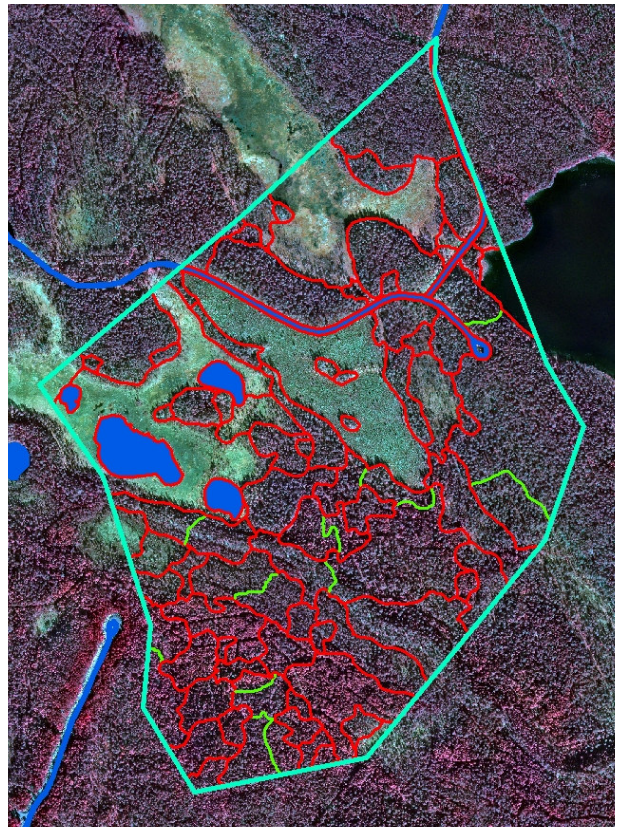

Figure 2. The segmentation result which is based on canopy height model and stand density estimates. Red lines are classified to be strong and green lines are weakly separating stands.

The result of specialist-made manual delineation is presented in Figure 3. Useful boundaries are colored red, while unnecessary were not found. The mean stand size in the resulting stand map was 1.9 ha.

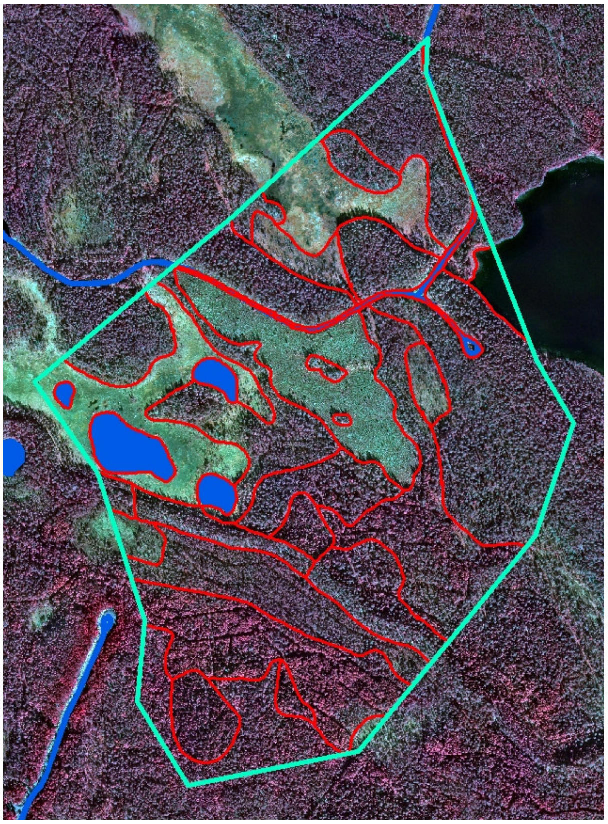

Figure 3. The segmentation result based on manual delineation of CIR imagery made by a professional forester.

Adjusted R^2's for segmentations are in table 2 and the line strengths in table 3. The differences between the three segmentations on total volume and pine volume (main species) are small. Spruce and deciduous tree volume separation benefited from CIR -supported segmentation (Table 2), which can be seen from the higher adjusted R^2's of segmentations.

Table 2. Comparison of adjusted R^2's on hierarchical segmentation approach; based on crown-height model (CHM), LIRGA segmentation approach; based on height, density and hardwood pixel percentage raster (H_D_HW) and manual segmentation based on CIR imagery (Manual).

Adjusted R^2	CHM	HDHW	Manual
G	0.64	**0.68**	0.66
DBH	**0.77**	0.71	**0.77**
H	**0.84**	0.80	0.83
V	0.66	**0.70**	0.65
N	0.76	0.75	**0.82**
Scots Pine			
G1	0.62	**0.69**	0.64
DBH1	**0.67**	0.66	0.66
H1	0.74	**0.75**	0.74
V1	0.66	**0.71**	0.66
N1	0.76	0.74	**0.80**
Norway Spruce			
G2	0.61	**0.70**	0.69
DBH2	0.51	**0.60**	0.54
H2	0.54	**0.63**	0.56
V2	0.65	**0.73**	0.68
N2	0.04	0.02	**0.14**
Hardwoods			
G3	0.32	**0.47**	0.41
DBH3	0.30	**0.39**	0.35
H3	0.31	**0.42**	0.36
V3	0.31	**0.50**	0.42
N3	0.46	0.33	**0.49**
Mean of $adjR^2$	0.56	**0.60**	0.59

Table 3. Comparison of mean line strengths weighted by length and percentages of unnecessary line in different delineation approaches.

	CHM	H_D_HW	Manual
Mean line strength	2.3	2.24	2.38
Total amount of line, meters	19570	20388	15252
% of unnecessary line	6	5	0

5. DISCUSSION

The results show that automatic approaches using LiDAR alone or a combination of LiDAR and CIR imagery can separate timber types with a good significance. While CIR imagery improves the separation of different species types, the LiDAR alone provides a good separation of timber size and density. This is in line with Packalen & Maltamo's (2007) results on the timber inventory: LiDAR alone can provide sufficient data for receiving stand totals, while using CIR band derivatives in the prediction increased the accuracy of the species-wise estimates.

The critical part in the stand delineation process is making the best possible choice for the stand definition. Traditionally, stand delineation has been an expert opinion combining many biological, ecological and operational issues. This is why it has been difficult to introduce any conclusive quality criteria for the result. However, to develop automatic approaches, well-defined criteria are necessary. The approach in this study has been to concentrate on delineating timber types apart, not worrying how they could be managed. In most cases in Scandinavian conditions, the timber types give also the basic units for operations. For economical reasons, the mean stand size, varying between 0.6ha to 0.8ha in the segmentations, is too small to be operated alone. An intelligent clustering process of timber stands, to receive operations units, has to be introduced on top of the timber stand delineation and attribution process. In this process, many additional layers of information have to be combined for high-quality decision-making. These layers could include for example digital terrain model information, soil type and soil operability, as well as hauling distance and hauling cost layer to reach the nearest road. Minimum sizes of operations need to be set for each operation type to include the economic constraints in the operations.

Some of the lines found by the delineation approaches considered in this analysis were deemed to be “unnecessary”. This is made based on the criteria analyzed, which is a simplification of the real life situations. It is possible some of the lines would be useful if judged based on other timber type criteria. However, the criteria used covers most of the typical timber typing situations in Scandinavia.

6. REFERENCES

Diedershagen, O., Koch, B., Weinacker, H., 2004. Automatic segmentation and characterisation of forest stand parameters using airborne LiDAR data, multispectral and FOGIS data. Proceedings of the ISPRS working group VIII/2 in Freiburg, Germany 03-06 October 2004.

Haara, A. and Haarala, M., 2002. Tree species classification using semi-automatic delineation of trees on aerial images. Scandinavian Journal of Forest Resources 17, pp. 556–565.

Hyyppä, J., Kelle, O., Lehikoinen, M., and Inkinen, M. 2001. A Segmentation-Based Method to Retrieve Stem Volume Estimates from 3-D Tree Height Models Produced by Laser Scanners IEEE Transactions on Geoscience and Remote Sensing, VOL. 39, NO. 5, MAY 2001 pp. 969-975

Leckie, D., Gougeon, F., Walsworth, N. and Paradine, D., 2003. Stand delineation and composition estimation using semiautomated indiviual tree crown analysis. Remote Sensing of Environment 85, pp. 355–369.

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. Volume 90, Issue 3, 15 April 2004, Pages 319-330.

Mustonen, J. 2007. Metsikkökuvioiden automaattinen segmentointi puuston latvustoa kuvaavan laserpintamallin ja ilmakuvan avulla. Pro Gradu, University of Helsinki. Depatrtnment of Forest Research Management.

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: 88–99.

Packalén, P. & Maltamo, M. 2007. The k-MSN method in the prediction of species specific stand attributes using airborne laser scanning and aerial photographs. Remote Sensing of Environment 109(3): 328–341.

Pekkarinen, A. 2002. Image segment-based spectral features in the estimation of timber volume. Remote Sensing of Environment 82(2-3): 349-359.

Pekkarinen, A. 2004. Image segmentation in multi-source forest inventory (PhD thesis). Metsäntutkimuslaitoksen tiedonantoja - The Finnish Forest Research Institute, Research Papers 926. 35 s. + 4 partial publications.

Pukkala, T. 2004. MONSU metsäsuunnitteluohjelmisto. Versio 4. Ohjelmiston toiminta ja käyttö. (Manual of MONSU forest management planning software) 75 s.

Sell, R. 2002. Segmentointinmenetelmien käyttökelpoisuus ennakkokuvioinnissa. Metsätieteellinen aikakauskirja 3/2002: 499-507.

Wang, M. Tseng, Y-H., 2004. LiDAR data segmentation and classification based on OCTREE structure. Geo-Imagery Bridging Continents XXth ISPRS Congress, 12-23 July 2004 Istanbul, Turkey. Congress Proceedings, Commission 3.

Frieke M.B. Van Coillie, Lieven P.C. Verbeke and Robert R. DeWulf. Semi-automated forest stand delineation using wavelet-based segmentation of very high resolution optical imagery in Flanders, Belgium.