Pukkala *Forest Ecosystems* (2020) 7:13
https://doi.org/10.1186/s40663-020-00221-8

Forest Ecosystems

RESEARCH Open Access

Delineating forest stands from grid data

Timo Pukkala

Abstract

Background: Forest inventories are increasingly based on airborne laser scanning (ALS). In Finland, the results of these inventories are calculated for small grid cells, 16 m by 16 m in size. Use of grid data in forest planning results in the additional requirement of aggregating management prescriptions into large enough continuous treatment units. This can be done before the planning calculations, using various segmentation techniques, or during the planning calculations, using spatial optimization. Forestry practice usually prefers reasonably permanent segments created before planning. These segments are expected to be homogeneous in terms of site properties, growing stock characteristics and treatments. Recent research has developed methods for partitioning grids of ALS inventory results into segments that are homogeneous in terms of site and growing stock characteristics. The current study extended previous methods so that also the similarity of treatments was considered in the segmentation process. The study also proposed methods to deal with biases that are likely to appear in the results when grid data are aggregated into large segments.

Methods: The analyses were conducted for two datasets, one from southern and the other from northern Finland. Cellular automaton (CA) was used to aggregate the grid cells into segments using site characteristics with (1) growing stock attributes interpreted from ALS data, (2) predicted cutting prescriptions and (3) both stand attributes cutting prescriptions. The CA was optimized for each segmentation task. A method based on virtual stands was used to correct systematic errors in variable estimates calculated for segments.

Results: The segmentation was rather similar in all cases. The result is not surprising since treatment prescriptions depend on stand attributes. The use of virtual stands decreased biases in growth prediction and in the areas of different fertility classes.

Conclusions: Automated stand delineation was not sensitive to the type of variables that were used in the process. Virtual stands are an easy method to decrease systematic errors in calculations.

Keywords: Cellular automata, Segmentation, Stand demarcation, Particle swarm optimization

Background

The use of airborne laser scanning (ALS) is increasing rapidly in forestry (Hyyppä et al. 2008; Vauhkonen et al. 2014). Airborne laser scanning can be done from airplanes or unmanned aerial vehicles. The scanning produces a cloud of return pulses, which can be used to interpret forest variables (Hyyppä et al. 2008). In a normal case, the x, y and z coordinates of the return points are recorded, as well as the intensity of the return pulse (Roncat et al. 2014). It is also known whether the pulse is reflected from ground or vegetation. Differences in the z coordinates between ground pulses and highest vegetation pulses constitute a canopy height model, which can be used to estimate tree locations and heights, and delineate tree crowns (Koch et al. 2014).

The tree-based interpretation approach uses the canopy height model to interpret tree locations and heights (Koch et al. 2014; Wing et al. 2018). Crown diameters can also be interpreted. Tree diameter can be predicted from tree height, crown diameter and other variables calculated from the pulse data. The pulse data can also be used to calculate variables that describe crown shape. These variables can be used to interpret tree species (Vauhkonen et al. 2009). The fact that small trees are often hided by

Correspondence: timo.pukkala@uef.fi
University of Eastern Finland, PO Box 111, 80101 Joensuu, Finland

the crowns of large trees has decreased the accuracy of forest inventories based on individual tree detection. However, new methods and algorithms are being developed which mitigate this problem (e.g., Kansanen et al. 2016).

The ALS data can also be interpreted with the area-based approach (Næsset 2014). In this method, the stand variables can be interpreted for any spatial units such as existing stand compartments, micro segments or grid cells. The method requires a set of field plots. The idea is to calculate the same variables from the ALS pulse data for the interpretation units and field plots. This makes it possible to find the most similar field plots for every interpretation unit. The measured stand variables of the most similar field plots are used to derive stand characteristics for the interpretation units. Also regression analysis can be used to predict the stand characteristics (Hyyppä et al. 2008). Some of the methods developed recently, such as distribution matching, cannot be easily classified to represent individual tree or area-based approach (Vauhkonen and Mehtätalo 2015; Kansanen et al. 2019).

In Finland, the area-based approach has been used to interpret forest variables for both existing stands and grid cells of 16 m by 16 m (http://www.metsaan.fi/paikkatietoaineistot). Both interpretation results cover all private forests of Finland and are freely available for forest planning. The use of stand-level interpretation results is straightforward and does not require new methodological development in forest planning (Pukkala 2019a).

The situation is different for the grid data. It is possible to treat the grid cells as calculation units in forest planning, and use spatial optimization to create continuous and large enough treatment units for the implementation of prescribed treatments (Heinonen and Pukkala 2007; Packalén et al. 2011; Pukkala 2019a). The spatial units created in this way are temporary, and the whole forest is never systematically divided into stand compartments. Due to their ephemeral nature, the treatment aggregations created by spatial optimization are often called as dynamic treatment units (Pukkala et al. 2014).

Dynamic treatment units may be a good approach when aiming at feasible treatment units and efficient utilization of forest resources (Heinonen et al. 2007). However, forestry practice often prefers fixed and reasonably permanent stand delineations. The prevailing ways to store, maintain and display forest data relies on stands (also called stand compartments or sub-compartments). Stands are also the basic units to simulate stand development and implement treatments.

Traditionally, the stand boundaries are visually demarcated on aerial photographs (Mustonen et al. 2008; Olofsson and Holmgren 2014). The stand boundaries are subjective and they may also be obsolete if new compartment inventories do not update stand borders. Stand variables interpreted from ALS data for grid cells offer possibilities to automatize, systemize and even optimize the stand delineation. An additional reason for using the ALS results interpreted for grid cells, instead of using variables interpreted for existing stands, is the fact that the results of area-based inventory are the most accurate if the area of the interpretation unit is approximately the same as the area of the field plots (Pascual et al. 2018). The interpretation grid of 16 m by 16 m results in 256-m^2 cells, which correspond to circular field plots of 9 m radius.

The stand compartments should be homogeneous in terms of site and growing stock characteristics, future stand development and treatments (Mustonen et al. 2008). In Finland, the most relevant site characteristics are soil type (e.g., peatland vs. mineral soil) and fertility class. It is especially important to delineate peatland forests and mineral soil forests into separate stands since they differ significantly in terms of accessibility and treatments. Mineral soil stands can be harvested all year round whereas peatland forest are accessible only when the soil is frozen. Site fertility dictates the growth rate or trees and is therefore an important variable if the stand delineation aims at homogeneous future development within the stand.

In addition to the homogeneity of site, growing stock and treatments, there are also other requirements for stand delineation. Stand shape should be “simple”, and irregular stand shapes are usually avoided. Stands should not be very small, but very large stands are also problematic. Stands are often regarded as undividable units, but too large stands might lead to the need to divide them into two or more parts when aiming for instance at the same harvested volume in every year.

Previous studies have developed methods to create stands or smaller segments from grid data (Baatz and Schäpe 2000; Mustonen et al. 2008; Wulder et al. 2008; Koch et al. 2009; Wu et al. 2013; Olofsson and Holmgren 2014; Pukkala 2019a, b). The suggested methods have used either various segmentation methods (e.g., Mustonen et al. 2008) or cellular automata (e.g., Pukkala 2019b). The segmentation or stand delineation can be done based on the ALS pulse data (Mustonen et al. 2008) or using variables interpreted for cells from the pulse data (Pukkala 2019a).

Similarity of treatment prescriptions has not been used in previous studies on automated, pre-planning stand delineation. Previous studies (e.g. Pukkala 2018) have shown that stand basal area and mean tree diameter correlate strongly with the financial maturity of the stand for cutting. The study of Kangas et al. (2011) suggests that errors in mean tree diameter may cause even greater inoptimality losses than errors in stand basal area. Mean diameter correlates with the relative value increment of trees, and therefore predicts the financial maturity of the trees for cutting.

Therefore, for the similarity of treatments within the stand it seems important to use stand basal area and mean tree diameter as the criteria of stand delineation, in addition to soil type and site fertility. However, also direct indicators for cutting maturity and cutting type have been developed. For example, Pukkala (2018) developed formulas that indicate the probability that cutting the stand immediately is the optimal decision. Other formulas were developed for the probability that partial cutting is optimal, thinning type and diameter at which thinning intensity is 50%. The predictions of these formulas can be used as additional criteria for stand delineation, to improve the results in terms of similarity of treatments.

The aim of the current study was to compare automated stand delineations based on (1) growing stock variables, (2) predicted treatments (3) or both, in addition to soil type and site fertility. The method used in stand delineation was cellular automaton (von Neuman 1966; Wolfram 2002) adapted from an earlier study (Pukkala 2019a). The data source was the ALS-based forest inventory where the stand variables are interpreted for grid cells. The data are available at http://www.metsaan.fi/paikkatietoaineistot. The study also proposed a new method to use grid data to deal with within-stand variation in planning calculations.

Materials

Site and growing stock variables interpreted for grid cells were used as the basis of stand delineation. The data are available at http://www.metsaan.fi/paikkatietoaineistot in packages of 375 × 375 cells. Since the size of the cell is 16 m × 16 m = 256 m^2, one package covers 3600 ha (375 × 375 × 256 m^2 = 36,000,000 m^2 = 3600 ha). Two randomly selected 3600-ha grids were downloaded from the web site, one representing southern Finland (x and y coordinates of the lower left corner 326,000 m and 6,756,000 m, respectively, in the GRS Transverse Mercator system) and the other representing northern Finland (590,000 m, 7,326,000 m).

The following variables (layers) were used for stand delineation: land category, soil type, fertility class, mean diameter, mean height, stand basal area. The growing stock variables are available for the total growing stock (all species combined) and separately for pine, spruce and broadleaf species. Variables interpreted for the total growing stock were used for stand delineation in this study.

From the fertility class layer and species-specific layers for stand basal area and mean diameter it can be concluded that both areas represented typical managed Finnish forests where. Many stands were mixtures of two or more tree species. Finnish forests are usually managed according to the principles of even-aged forestry and younger stands are often plantations. Despite this, the stands could have large within-stand variation in tree diameter, due to the gradual advance regeneration of various species, especially spruce. In the southern case study forest, mesic sites were the most common, followed by sub-xeric sites. In the north, subxeric and xeric sites covered most of the area.

The land category layer indicates whether the cell represents productive forest, stunted forest, waste land, agricultural land, water, road, etc. This layer was used as a mask to filter out all cells that did not represent productive forest. The area of productive forest was 2977 ha in the southern case study area and 2140 ha in the northern area. Lakes, nonproductive peatlands etc., fragmented the productive forest clearly more in the north. All the other layers (two site variables and three growing stock variables) were used in stand delineation. The soil type layer indicates whether the cell belongs mineral soil or peatland, which is subdivided into spruce mire, pine bog and open bog (no trees). The fertility class indicates whether the cell is mesotrophic, herb rich, mesic, subxeric, xeric or barren heath site.

Methods

Treatment variables

The site and growing stock variables available for the grid cells were used to calculate additional variables for each cell. These variables describe the probability of cutting (*pCut*), probability of thinning (*pThin*), thinning type (*tType*) and diameter at which thinning intensity is 50% (*d*50). These additional variables were calculated with the models of Pukkala (2018). The first model predicts the probability that cutting the stand immediately is the optimal decision. The second model predicts the probability that partial cutting is optimal in case of cutting. The model for thinning type indicates the degree to which the optimal partial cutting is thinning from above. The fourth model gives the diameter class for which the thinning intensity should be 50%. In thinning from above, most harvested trees are larger than this diameter whereas the opposite is true in thinning from below.

The equations for the cutting variables are as follows (Pukkala 2018):

$$pCut = \frac{1}{1 + \exp[-f(\mathbf{x})]} \quad (1)$$

(with)

$$\begin{aligned} f(\mathbf{x}) = {} & 16.032 - 1.098\sqrt{G} + 2.806 \ln D \\ & \times \ln G - 0.573\sqrt{D} \times \sqrt{D} + 0.000088G \times TS \\ & + 0.00454G \times R + 0.000486TS \\ & \times \ln R - 3.878 \ln R - 0.944FS - 0.814VT \end{aligned}$$

$$pThin = \frac{1}{1 + \exp\left[-\left(-12.208 + 10.103 \ln G - 4.004\sqrt{G} + 0.00731D \times R + 0.856\sqrt{G_{\text{pulp}} - 0.785FS}\right)\right]} \quad (2)$$

$$\begin{aligned} d50 = {} & 11.324 - 0.440G - 4.360 \ln D \\ & + 0.597\sqrt{G} \times \sqrt{D} + 0.000832TS \\ & \times \ln R + 0.000170G \times TS - 0.000117D \\ & \times TS + 0.175 \ln G_{\text{pulp}} - 0.0126D \times R \end{aligned} \quad (3)$$

$$\begin{aligned} tType = {} & (0.980 + 0.964\sqrt{D} - 0.000619D \times G - 0.0630 \ln D \\ & \times \ln G)^2 + 0.423^2 \end{aligned} \quad (4)$$

where D is basal-area-weighted mean diameter of trees (cm), G is stand basal area ($m^2 \cdot ha^{-1}$), G_{pulp} is the basal area of pulpwood-sized trees (dbh 8–18 cm) in $m^2 \cdot ha^{-1}$, TS is temperature sum (degree days > 5 °C), R is discount rate (%), FS is indicator variable for fertile growing sites (mesic or better) and VT is indicator variable for sub-xeric site.

The above cutting variables depend on mean tree diameter, stand basal area, fertility class, temperature sum and discount rate. The temperature sum of the region was obtained from a temperature sum map (https://ilmatieteenlaitos.fi/terminen-kasvukausi), and the discount rate was taken as 3%. All other predictors of the cutting models were obtained from the ALS-based inventory results interpreted for the grid cells.

Cellular automaton

The cellular automaton (CA) of Pukkala (2019a) was adapted for the purposes of the current study. The automaton was modified so that the four additional cutting variables could also be used in stand delineation. The used CA first divides the forest into square-shaped initial stands (3 ha in this study). Then, the "most suitable" stand number is given to each cell for several iterations. A cell gets the stand number of one of its adjacent stands, which are eight in number (stand number of the cell to the east, west, north and south, and four corner cells). Figure 1 illustrates the evolution of the stand delineation during a short CA run.

The stand number given to a cell depends on a score that is calculated to its eight neighbor cells:

$$S_{ij} = a_1 s_1(B_{ij}) + a_2 s_2(A_j) + a_3 s_3(D_{ij}) \quad (5)$$

where S_{ij} is the score of assigning the number of stand j to cell i, B_{ij} is the proportion of common border between cell i and stand j, A_j is the area of stand j, and D_{ij} is the difference in stand characteristics between cell i and stand j. Increasing stand area and common border increase the score, and increasing difference in stand variables between the cell and the stand decreases the score. If the weight of the difference in stand characteristics (a_3) is high, the stands tend to be small and irregular with little within-stand variation in stand variables whereas high weight for common border (a_1) results in compact and round-shaped stands.

The area of the final stands is controlled through weight a_2 and function s_2, which indicate the effect of stand area on the score. If the purpose of stand delineation is to avoid both small and very large stands, a sigmoid type of function is preferred (Fig. 2). The same applies for function s_1, which describes the effect of the proportion of common border between cell i and stand j on the score of stand j (Fig. 2). The "proportion of common border" is the average of the weights of the 8 neighboring cells in such a way that the cell to east, west,

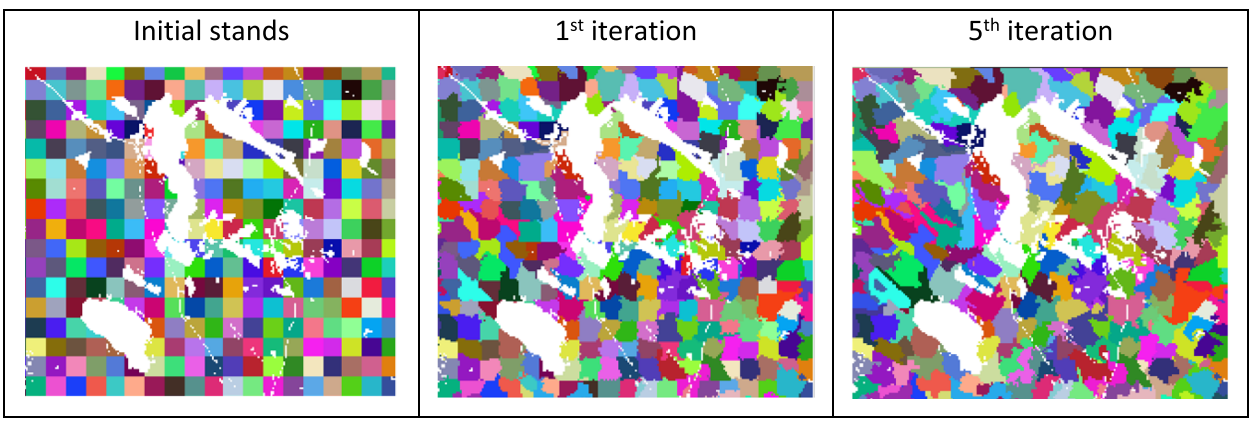

Fig. 1 Development of stands during five iterations of the cellular automaton used in this study. The size of the initial stands (left) is 3 ha (exactly: 3.0976 ha = 121 cells). Each cell is joined to one of its adjacent stands for several iterations

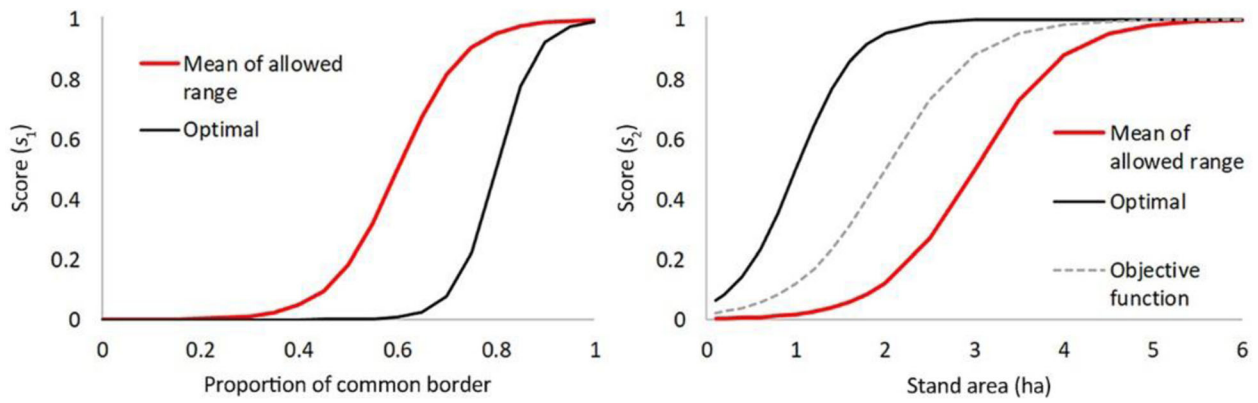

Fig. 2 Shapes of the functions that describe the effect of common border between a cell and a stand (left) and the effect of stand area (right) on the probability that the cell is joined to the stand

north and south gets a weight equal to one and a corner cell gets a weight smaller than one.

The functions that describe the effect of the proportion of common border (s_1) and stand area (s_2) on the score were as follows:

$$s_1 = \frac{1}{1 + \exp\left(b_1\left(B_{ij} - b_2\right)\right)} \tag{6}$$

$$s_2 = \frac{1}{1 + \exp\left(c_1\left(A_j - c_2\right)\right)} \tag{7}$$

The difference between stand characteristics in cell i and stand j was calculated from:

$$D_{ij} = \sqrt{\sum_{n=1}^{N} w_n \left(q_{ni} - q_{nj}\right)^2} \tag{8}$$

where w_n is the weight of stand variable n ($\Sigma w_n = 1$), q_{ni} is the value of standardized variable n in cell i and q_{nj} is the mean value of the same variable in stand j.

In this study, the number of variables used to calculate the difference (N) was 5, 6 or 9 (two site variables and 3, 4 or 7 other variables). When the delineation was based on stand variables (case 1), the growing stock variables used to measure the similarity of the cell and an adjacent stand were mean tree diameter, mean tree height and stand basal area. When the delineation was based on forecasted cuttings these variables were replaced by probability of cutting, probability of thinning, thinning type and diameter at which thinning intensity is 50% (case 2). When the delineation was based on both growing stock variables and treatments (case 3), all the above variables were used to measure similarity. The two site variables used in all cases were soil type and fertility class. All variables included in the calculations were standardized to mean zero and standard deviation 1. This removed the effect of different units of variables.

Functions 6 and 7 have two parameters. In Fig. 2, the red curves show the effect of stand area and common border with the mean values of the allowed ranges of the parameters and the black curves show the effect with the most common values when the parameters were optimized for different stand delineation tasks. The dashed curve on the right-hand-side diagram describes the effect of average stand area on the objective function when the parameters of the cellular automaton were optimized (see Eq. 10 below).

A stand might disappear completely during a CA run. A stand may also become divided into two or more non-connected parts. Since this was not a wanted outcome, non-connected parts were given unique stand numbers after every five iterations. Therefore, the number of stands may decrease or increase during a CA run. The total number of iterations in one CA run was 17 (Pukkala 2019a).

Optimizing the cellular automaton

The performance of the CA depends on its parameters. The comparison of different stand delineation cases is the most objective if the parameters of the CA are optimized for each stand delineation case. The parameters affecting the performance of the CA were as follows (allowed ranges shown in square brackets):

- Weight of the corner cell in the calculation of common border [0, 0.3]
- Weight of common border in Eq. 5 (a_1) [0.4, 0.7]
- Weight of stand area in Eq. 5 (a_2) [0.1, 0.4]
- Weight of similarity of stand variables in Eq. 5 (a_3) [0.2, 0.5]
- Weights of 5, 6 or 9 stand variables in Eq. 8
- Two parameters of Eq. 6 that describes the effect of common border (s_1)
- Two parameters of Eq. 7 that describes the effect of stand area (s_2)

When five or six variables were used to measure the similarity of stand variables in the cell and an adjacent stand, the weights of these variables (w_n in Eq. 8) were allowed to range from 0.1 to 0.6. When similarity was measured with nine variables, the allowed range of each weight was 0.05–0.5. The allowed ranges of parameters b_1, b_2, c_1 and c_2 (Eqs. 6 and 7) were [– 25, – 5], [0.4, 0.8], [– 3, – 1] and [1, 5], respectively.

The total number of CA parameters was either 13 (case 1), 14 (case 2) or 17 (case 3). Their optimal values were found by using particle swarm optimization in the same way as explained in detail in Pukkala (2019b). The baseline objective variable maximized in optimization was the average degree of variance explained by the stand delineation. The $R2$ of a certain variable was calculated from:

$$R2 = 1 - SSE/SST \tag{9}$$

where SSE is the sum of squared deviations from stand mean and SST is the sum of squared deviations from the overall mean. However, this baseline objective variable was modified by the mean area of the delineated stands and the proportion of very small stands. The modified objective function was as follows:

$$\text{Max } z = 0.8\overline{R2} + 0.2s_2\left(\overline{A}\right) + 0.2P_{\text{small}} \tag{10}$$

where $\overline{R2}$ is the average degree of explained variance of the 5, 6 or 9 stand variables used in delineation, $\overline{A}$ is the average area of the obtained stands, s_2 is a function that describes the effect of $\overline{A}$ on objective function (Fig. 2, dashed line), and P_{small} is the proportion of stands smaller than 0.1 ha (i.e. smaller than 4 cells since one cell is 0.0256 ha). The weights of the criteria were set subjectively based on preliminary optimizations.

If the degree of explained variance ($R2$) is maximized without any constraints or other objectives, the CA produces delineations consisting of a high number of small and irregularly shaped stands. Since this kind of delineation would be hardly acceptable for forestry practice, parameter optimization was controlled by restricting the ranges of the parameters and adding the mean stand area and the proportion of small stands to the objective function (Eq. 10). The parameters of the CA were optimized separately for all stand delineation tasks, i.e., six times (cases 1, 2 and 3 for southern and northern Finland).

Results

Optimal CA-parameters

The optimal weight of the corner cell was equal or close to its maximum allowed value (0.3) in most optimization cases (Table 1). Of the three criteria that determined the stand to which a cell was joined, similarity of stand variables had the smallest weight in northern Finland while stand area had the smallest weight in the south. The weight of the proportion of common border was often equal to its lowest allowed value (0.4) because this variable, which contributes to round and compact stand shape, contradicts with the criteria of the objective function that was maximized in parameter optimization (Eq. 10). In the north, much weight was given to stand area because it was otherwise difficult to reach large mean stand area in the fragmented forest landscape of the northern case study area. Productive forest of the northern area was more fragmented by lakes, wasteland (e.g. unproductive peatland) etc., which made it more difficult to create large and homogeneous stands in the north.

Of the stand variables used for segmentation, mean height had often the lowest allowed weight, or when the weight of mean height was high (case 3 in northern Finland), the weight of mean diameter was low. This result suggests that it is sufficient to use either mean diameter or mean height as a criterion in automated stand delineation.

Of the variables that describe cutting treatments, the thinning type variable had always the lowest allowed weight, suggesting that this variable might not be needed in stand delineation. This variable describes the degree to which a partial cutting is thinning from above. The study of Pukkala (2018) suggests that the optimal thinning is usually thinning from above, implying that there might be little variation in the optimal thinning type, making it unnecessary to use it in stand delineation.

In northern Finland, the weights of all cutting variables were equal to their smallest allowed values when the delineation was based on both growing stock cutting variables (case 3). The result suggests that cutting variables may not be important for stand delineation in northern Finland when maximizing Eq. 10. The parameters of the functions for the effect of stand area and proportion of common border (b_1, b_2, c_1, c_2) were also often on the limit of the allowed range (black curves in Fig. 2).

Degree of explained variation

The degree of explained variance of stand variables was high, 0.777–0.953, for soil type and reasonably high (0.695–0.893) also for fertility class (Table 2). When the stand delineation was based on growing stock variables (cases 1 and 3) the $R2$ of these variables was, with one exception, higher than in the case (case 2) where basal area, mean diameter and mean height were not used in delineation. The degree of explained variance in cutting variables was usually lower when these variables were not used in stand delineation. The $R2$ of a certain variable correlated positively but weakly with the weight of the variable (w_n in Eq. 8) in stand delineation (Fig. 3), suggesting that if the stands need to be homogeneous in

Table 1 Optimal parameter values of the cellular automaton used for stand delineation. Similarity of cells and stands was measured using growing stock variables (Case 1), forecasted cuttings (Case 2) or both (Case 3)

	South Finland			North Finland		
	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
Weight of corner cell	0.300	0.299	0.220	0.215	0.278	0.090
a_1 (common border)	0.400	0.514	0.480	0.400	0.400	0.401
a_2 (stand area)	0.300	0.191	0.256	0.400	0.350	0.396
a_3 (stand variables)	0.300	0.294	0.264	0.200	0.250	0.203
w_1 (soil type)	0.169	0.288	0.132	0.260	0.208	0.220
w_2 (fertility)	0.144	0.254	0.123	0.205	0.165	0.170
w_3 (mean diameter)	0.296	–	0.120	0.168	–	0.122
w_4 (mean height)	0.095	–	0.092	0.100	–	0.197
w_5 (basal area)	0.296	–	0.120	0.266	–	0.091
$w_{3/6}$ (cut probability)	–	0.169	0.120	–	0.176	0.050
$w_{4/7}$ (thin probability)	–	0.101	0.120	–	0.176	0.050
$w_{5/8}$ (thin type)	–	0.091	0.050	–	0.100	0.050
$w_{6/9}$ (thin diameter)	–	0.091	0.120	–	0.176	0.050
b_1 (border effect)	−16.946	−25.000	−25.000	−5.000	−11.716	−5.000
b_2 (border effect)	0.800	0.800	0.800	0.800	0.800	0.800
c_1 (area effect)	−3.000	−3.000	−3.000	−1.000	− 3.000	− 1.000
c_2 (area effect)	1.000	1.000	1.000	1.000	1.021	1.000

terms of a particular variable, this variable should have a high weight in stand delineation.

In the southern case study area, the average $R2$ of all nine stand variables was the highest when both growing stock and cutting variables were used to measure the similarity of a cell and an adjacent stand (Table 2). In northern Finland, the highest overall $R2$ was obtained when the stand delineation did not use growing stock variables. However, a high degree of explained variance was associated with small average stand area and high proportion of small stands. If the quality of the delineation is measured with the objective function used in parameter optimization (Eq. 10), it can be concluded that the quality of delineation was the lowest when stand variables (basal area, mean diameter, mean height) were not used as criteria (case 2, lowest line in Table 2). The other cases, namely using only growing stock variables alone or with cutting variables resulted in rather similar objective function values.

Similarity of stand delineations

Visual inspection of the stand boundaries reveals that different CA runs resulted in fairly similar stand boundaries (Figs. 4 and 5). Especially cases 1 and 3, where the delineation employed growing stock variables with or without cutting variables, produced stands that were often practically identical (yellow and red borders in Figs. 4 and 5). Stand boundaries that were based on predicted cutting variables (light blue borders in Figs. 4 and 5) differed more from the other delineations but also in this case many boundaries coincided with the boundaries of the other delineations. In southern Finland, the average stand area was clearly larger and the proportion of small stands lower when only cutting variables were used. The opposite was true for northern Finland.

Complications

Previous research and theoretical reasoning (Pukkala 1990; Pukkala and Kolström 1991) have shown that increasing within-stand variation in stand density and tree size decreases the volume increment of the stand, compared to homogeneous stands with the same average stand basal area and mean diameter. When the growing stock is described only with the average stand balsa area and mean diameter, it can be expected that growth predictors will be overestimated, the most in the most heterogeneous stands.

Another consequence of condensing the cell level information to a single stand level record is the underestimation of the areas of rare fertility classes (Pukkala 2019a). In a normal case, the most common fertility class of the cells that constitute a stand is given to the whole stand. Underestimated areas of rare sites are consequences of the fact that these sites often occur as small spots within more common fertility classes. If large stands are pursued, these small spots are not demarcated as separate stands.

Table 2 Degree of variance explained by the stand delineation (between-stand variation/total variation) and other statistics for the stand delineation. *R2* of variables used in stand delineation are in italics. The values of the criterion variables used in parameter optimization are shown in boldface. The objective function value of the optimization is also in boldface

	South Finland			North Finland		
	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
Soil type, *R2*	*0.819*	*0.865*	*0.777*	*0.953*	*0.926*	*0.953*
Fertility class, *R2*	*0.710*	*0.752*	*0.695*	*0.893*	*0.840*	*0.876*
Mean diameter, *R2*	*0.682*	0.516	*0.633*	*0.553*	0.512	*0.574*
Mean height, *R2*	*0.729*	0.566	*0.683*	*0.621*	0.604	*0.642*
Basal area, *R2*	*0.615*	0.574	*0.634*	*0.442*	0.501	*0.442*
Cutting probability, *R2*	0.615	*0.578*	*0.643*	0.452	*0.539*	*0.460*
Thinning probability, *R2*	0.565	*0.590*	*0.666*	0.354	*0.569*	*0.388*
Thinning type, *R2*	0.499	*0.535*	*0.576*	0.302	*0.524*	*0.337*
Thinning diameter, *R2*	0.588	*0.537*	*0.671*	0.357	*0.574*	*0.399*
Average *R2*	**0.647**	**0.613**	**0.664**	**0.547**	**0.621**	**0.563**
Minimum area (ha)	0.03	0.03	0.03	0.03	0.03	0.03
Mean area (ha)	**3.51**	**4.36**	**3.44**	**1.95**	**1.43**	**1.83**
Maximum area (ha)	13.98	20.71	15.87	15.08	17.20	14.13
Small stands (< 0.1 ha) (%)	**11.9**	**8.5**	**13.2**	**22.3**	**33.0**	**25.0**
Number of stands	829	685	868	1099	1493	1168
Objective function	**0.684**	**0.671**	**0.694**	**0.488**	**0.479**	**0.483**

In the southern case study forest, the proportion of herb rich site was underestimated by 26%, and in the northern forest, the proportion of mesic site was underestimated by 9%, when compared to proportions calculated from cells (Table 3). The "missing area" of rare fertility classes was moved to the areas of more common adjacent fertility classes. The overestimate of predicted volume increment, calculated by the models of Pukkala et al. (2013), was 2.5% for southern Finland and 2.7% for northern Finland.

Remedies

The availability of cell level information makes it possible to calculate improved estimates for the areas of fertility classes, volume increment, and other variables. A method titled "virtual stands" was developed and

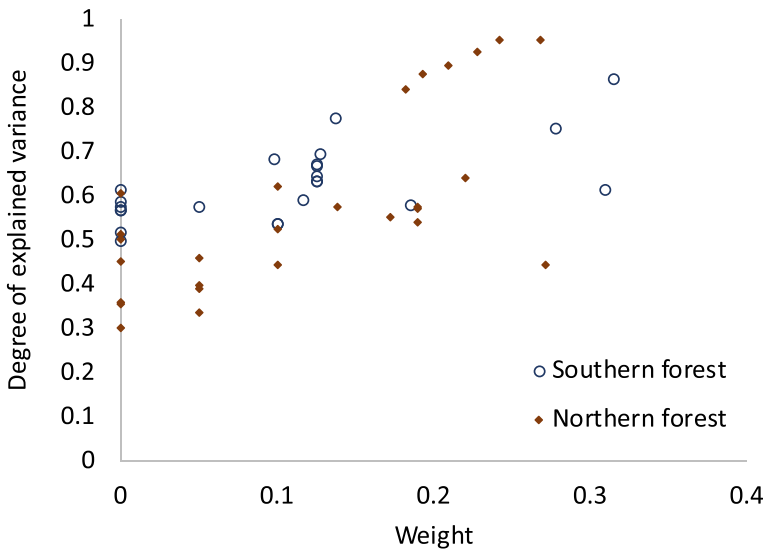

Fig. 3 Correlation between the weight of a stand variable (w_n in Eq. 8) and proportion of variance of the same variable explained by the stand delineation

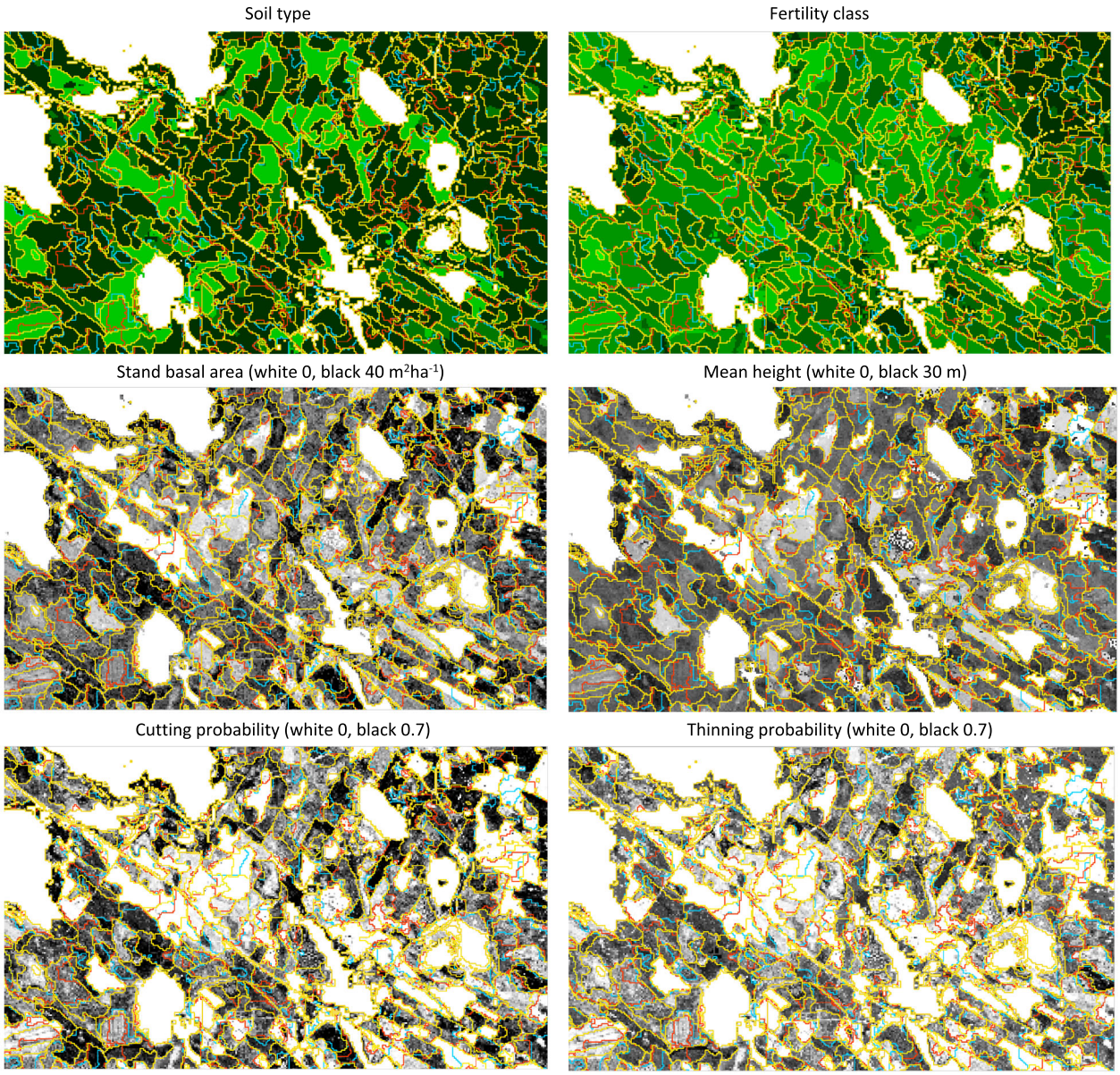

Fig. 4 Stand boundaries of the southern case study area obtained from cellular automaton with site, growing stock and cutting variables. The soil types from darkest to lightest shading are mineral soil, spruce mire and pine bog. In the fertility map, decreasing darkness of the shading implies decreasing fertility. Yellow stand boundaries are based of growing stock variables (case 1), blue boundaries on cutting variables (case 2) and red boundaries on all variables (case 3). The map shows about 900 ha of the total area of 3600 ha

tested in the current study. In this method, additional variables were calculated from the cell-level values and these additional variables were used to split the delineated stands into several stand records (virtual stands) that reflected within-stand variation in site productivity and growing stock variables. These additional variables were: proportions of different site fertility classes, standard deviation of stand basal area, standard deviation of mean diameter, and correlation coefficient between stand basal area and mean diameter.

The additional variables were used to create virtual stands as follows. First, the proportions of different fertility classes were used to divide the stand into several sub-stands, their areas corresponding to the areas of fertility classes. Second, each sub-stand was partitioned into five virtual stands that differed in basal area and mean diameter. It was assumed that the two growing stock variables were multi-normally distributed. Basal area was deviated from its average value by one standard deviation and the mean diameter corresponding to the deviated basal area

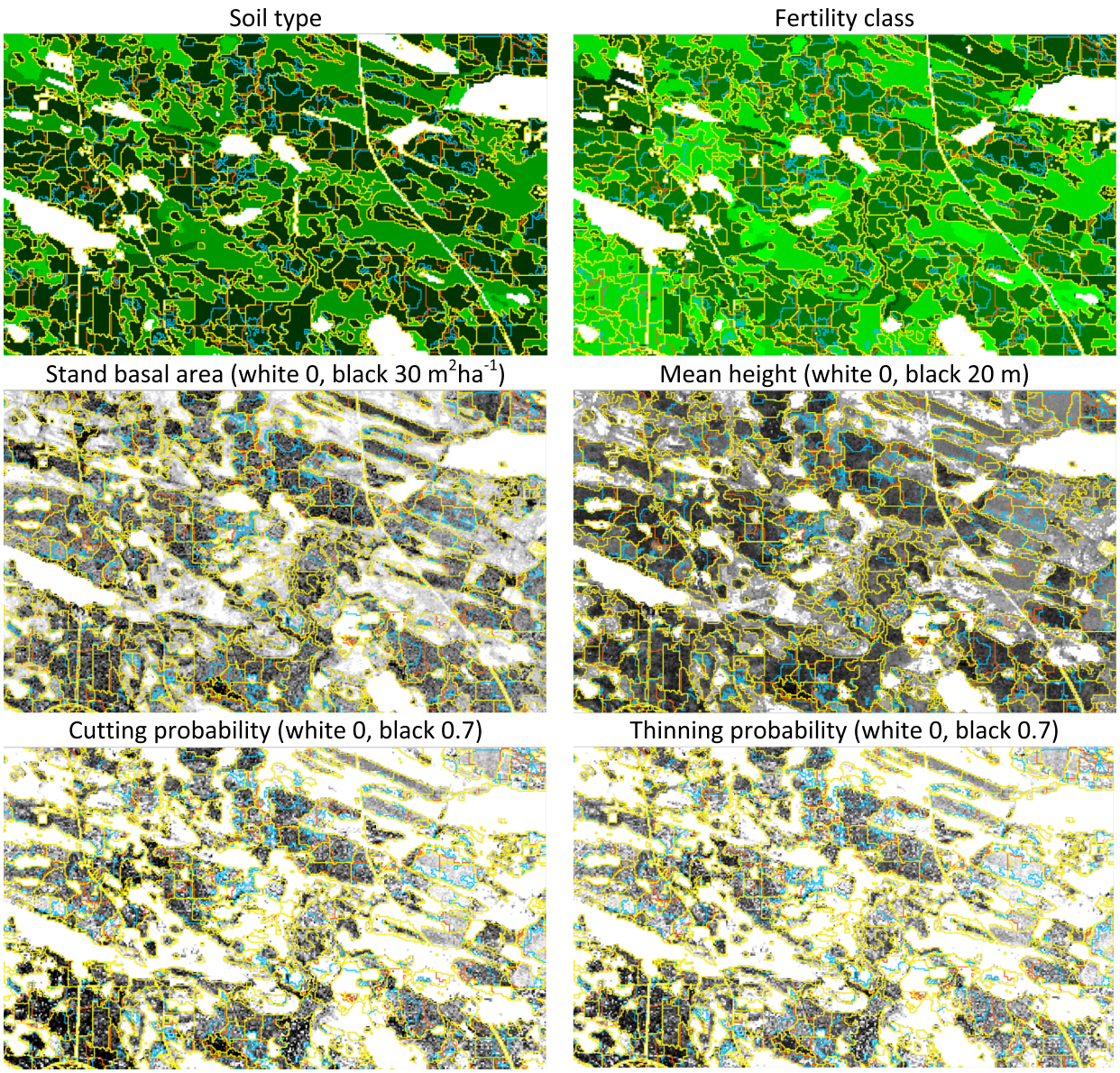

Fig. 5 Stand boundaries of the northern case study area obtained from cellular automaton with site, growing stock and cutting variables. The soil types from darkest to lightest shading are mineral soil, spruce mire, pine bog and open bog. In the fertility map, decreasing darkness of the shading implies decreasing fertility. Yellow stand boundaries are based of growing stock variables (case 1), blue boundaries on cutting variables (case 2) and red boundaries on all variables (case 3). The map shows about 900 ha of the total area of 3600 ha

was calculated using the standard deviations and the correlation coefficient. Then, mean diameter was deviated in the same way and the stand basal area was calculated for the deviated mean diameters (Fig. 6). The mean tree heights corresponding to deviated mean diameters were calculated with a height model (Siipilehto 1999), which was scaled so that the model prediction was equal to average mean height of the cells when diameter was equal to average mean diameter of the cells.

Together with average values of mean diameter and stand basal area the deviated values resulted in five combinations of stand basal area and mean diameter. The proportions of these combinations were obtained from the probability density function of the multinormal distribution. The area of the sub-stand was divided among the five virtual stands according to these proportions.

Table 3 shows that the use of virtual stands removed the biases from the proportions of different fertility classes. The bias of the growth prediction was also greatly reduced, especially in southern Finland. In northern Finland, the bias was reduced only about 60%. One

Table 3 Proportions of fertility classes and predicted mean annual volume increment when calculated from the characteristics of cells, stands or virtual stands

	South Finland			North Finland		
	Cells	Stands	Virtual stands	Cells	Stands	Virtual stands
Herb rich (%)	14.1	10.4	14.1	0.2	0.2	0.2
Mesic (%)	46.4	49.9	46.5	13.0	11.8	13.0
Sub-xeric (%)	32.3	33.2	32.3	57.2	58.3	57.3
Xeric (%)	7.1	6.4	7.0	29.5	29.7	29.6
Volume increment ($m^3 \cdot ha^{-1} \cdot a^{-1}$)	6.76	6.94	6.80	2.93	3.01	2.96
Bias (%)	–	2.5	0.3	–	2.7	1.1

reason for the remaining bias might be that stand basal area and mean tree diameter correlate with site fertility, which was ignored in the creation of virtual stands.

Discussion

ALS grid data are available from all private forests of Finland but this free data source is rarely used in forest planning now (2020). The reason is the lack of methodologies and experience to deal with datasets that differ from traditional data formats. The situation is similar also in other countries: more fine-grained inventory data, compared to traditional stand-level inventories, are becoming available for forest planning with the increased use of ALS, but forest planners are not prepared to fully utilize the potentials of the new data formats.

The way in which ALS data were used in the current study is only one possibility among several alternatives. As an alternative approach, the ALS pulse data could be used more directly for stand delineation instead of using stand variables interpreted from pulse data for cells (Mustonen et al. 2008; Olofsson and Holmgren 2014). Variables that describe the pulse data at certain location (in a very small cell or polygon) can be aggregated into homogeneous areas, often called as micro segments (Pascual et al. 2018). Then, the stand variables are interpreted for these segments instead of grid cells. This approach would eliminate the problem of mixed cells, which are cases where different parts of the cell belong to different stands, i.e. the cell is located at stand boundary (Pascual et al. 2018).

The third approach is to interpret individual trees from the pulse data and use the trees in planning calculations (Wing et al. 2018), or calculate the stand variables required in panning from the predicted

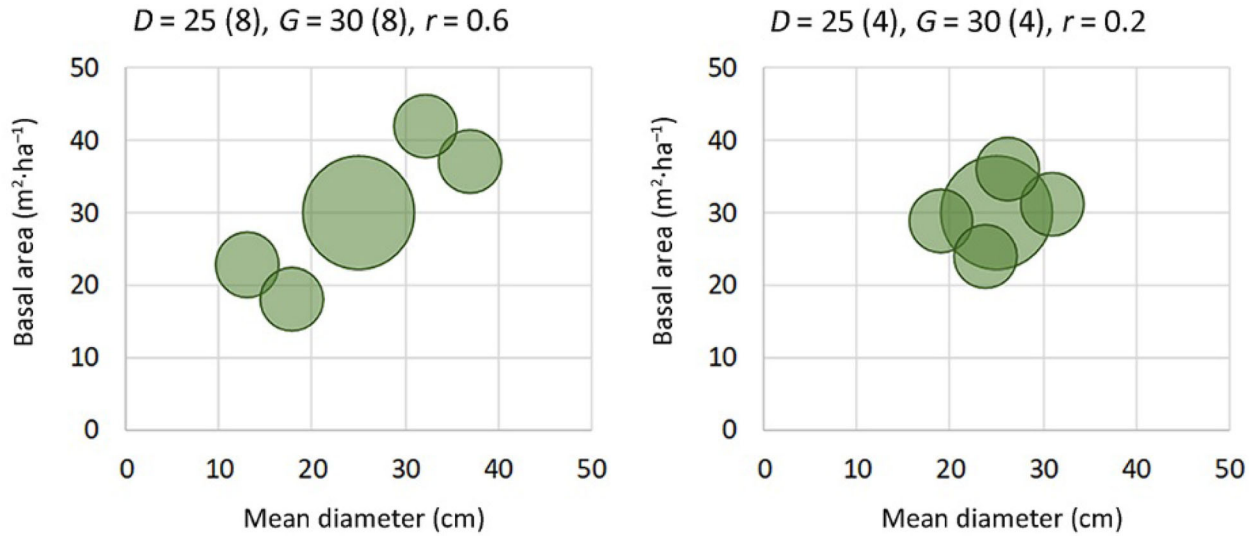

Fig. 6 Basal areas and mean diameters of the five virtual stands generated for a sub-stand based on the mean and standard deviation of stand basal area and mean diameter and their correlation. The size of the circle represents the relative area of each virtual stand, calculated from multi-normal probability density function. In both cases, the average value of mean diameters of the cells that constitute the stand (*D*) is 25 cm and the average stand basal area of the cells (*G*) is 30 $m^2 \cdot ha^{-1}$. The standard deviations (in brackets) are larger on the left-hand-side diagram and the correlation between basal area and mean diameter is stronger

characteristics of individual trees (Hyyppä et al. 2008). All these alternative lines to use ALS data in forest planning require further research and new methodological development.

This study alleviated the problem of lacking experience in the use of ALS data by developing methodologies for employing ALS-based grid data in forest planning. A cellular automaton was used to aggregate grid cells into stands that correspond to traditional stand compartments. As an enhancement to previous research, also the forecasted cuttings were used as criteria in stand delineation. Another enhancement was a new method, based on virtual stands, to deal with within-stand variation in planning calculations.

When delineating stands from the grid data, much weight was given to the stand area criterion. The purpose was to have stands that are large enough to serve as treatment units in the implementation of the plan. Large stands make it unnecessary to use spatial optimization to aggregate treatments. Spatial optimization has been used for a long time in forest research (e.g., Borges and Hogansson 1999; Lu and Eriksson 2000; Bettinger et al. 2002; Jumppanen et al. 2003) but its use in practical forest planning in Finland is very limited. Therefore, the methods developed in this study for stand delineation are easily adaptable to the current forest management systems.

The other methodological development, namely the use of virtual stands, requires that a few more variables are imported and stored in the forest database (the new variables are standard deviation of basal area and mean diameter, correlation coefficient between basal area and mean diameter, and proportions of fertility classes). Another required modification is the generation of the virtual stands as the first step of calculation, which is not a difficult step.

Reasonably large stands and avoidance of small stands are usually wanted in forest planning because this eliminates the need to use spatial optimization in planning calculations. However, large stand size increases within-stand variation in stand variables and decreases the proportion of variance explained by the stand delineation (Mustonen et al. 2008). For example, in the study of Pukkala (2019a) the $R2$ of mean diameter was 0.814 (within stand standard deviation, s_D, 0.61 cm) when the average stand area was as small as 0.36 ha. The $R2$ decreased to 0.692 (s_D, 0.73 cm) when the average stand area was 1.27 ha. In the current study, the $R2$ of mean diameter was 0.682 (s_D, 0.43 cm) in the southern case study forest) with mean stand area of 3.51 ha, and 0.516 (s_D, 0.55 cm) with 4.36 ha mean stand area. Therefore, aiming at large stands in automated stand delineation increases the need to consider within-stand variation in planning calculations.

If within-stand variation is ignored, the use of large stands results in overestimated growth prediction and underestimated areas of rare site types (Pukkala 2019a). The same problem of course exists with the traditional, visually demarcated stands (Pukkala 1990). Sometimes, other errors may cancel the bias. For example, the used growth models may underestimate growth. However, using information on within-stand variation helps to remove one source of error from growth predictions. The other sources of error must be corrected using other methods.

Within-stand variation may also affect the optimal treatments. For example, Pukkala and Miina (2005) found that heterogeneous stands should be thinned at lower average basal area than homogeneous stands. Therefore, there might be a need to use the virtual stands also in the simulation of treatment alternatives for stands. This would improve the predictions of stand development, optimality of treatment prescriptions, and estimated timber removals and incomes.

The results of this study showed that automated stand delineation with a cellular automaton produces stand boundaries that closely follow the borders between mineral soil and peat land. The degree of explained variation is also quite high for fertility class, but a part of the stands might be mixtures of two or more fertility classes. Some very small stands remained in all delineations (10%–20% of the number of stands). This outcome is partly unavoidable since Finnish landscapes often have small isolated patches of productive forest within agricultural lands or as islands in lakes and nonproductive peatlands.

The results also showed that the stand delineation is not very sensitive to variables used in the cellular automaton. However, the optimal weights of the used variables suggest that stand basal area should be used, together with either mean height or mean diameter. For the avoidance of inoptimality losses due to non-optimal prescriptions, mean diameter has been found to be even more important than stand basal area (Kangas et al. 2011). Of the cutting variables, the calculated probability of cutting seems to be the most important for good stand delineation. Thinning probability and thinning type variables are less important, most probably because these variables do not vary much. According to Pukkala (2018), the optimal cutting is almost always thinning, and the optimal type of thinning is almost always thinning from above. The thinning diameter (diameter at which thinning intensity is 50%) correlates strongly with mean diameter, and is therefore not necessary as a separate criterion for stand demarcation. Therefore, the use of the following variables can be recommended for automated stand delineation: soil type, fertility class, stand basal area, mean diameter, and probability of cutting.

The calculated probability that immediate cutting is the optimal decision depends on mean diameter, stand basal area, fertility class and temperature sum, in addition to discount rate (Eq. 1). The variable can be interpreted to describe the interactions of these variables.

Parameter optimization is not necessary in the practical application of the cellular automaton. Based on the results reported in this article and a few additional CA runs, the following parameter values can be recommended: the weights of common border, stand area and similarity of stand variables (a_1, a_2 and a_3 in Eq. 5) may all be 0.333; the weights of soil type, fertility class, mean diameter, stand basal area and cutting probability (Eq. 8) should be 0.3, 0.2, 0.2, 0.2 and 0.1, respectively (the weights of the other variables should be zero); parameters b_1, b_2, c_1 and c_2 of Eqs. 6 and 7 should be − 25, 0.8, − 3 and 1, respectively; and the weight of the corner cell should be 0.3.

When the cellular automaton was optimized in this study, only two of the three criteria that were judged important for good stand delineation, namely small within-stand variation and large enough stand area, were included in the objective function that was maximized in particle swarm optimization. The third criterion, stand shape, was pursued by constraining some of the CA parameters within certain limits. For example, the maximum weight of corner cell was 0.3 and the minimum weight of common border (Eq. 1) was 0.4. The manner in which the common border affected the probability of joining a cell to a certain stand (Fig. 2) was also constrained. These constraints prevented the CA from creating irregular (e.g. very long and narrow) stands.

Future research may develop methods to include also stand shape in the objective function. One possibility would be to use the area-perimeter ratio or the form heterogeneity variables suggested by Baatz and Schäpe (2000). Other segmentation methods and spatial optimization should also be tested in automated stand delineation. In addition, different ways to deal with within-stand variation in forest planning should be developed and compared. Alternative ways to create virtual stands and use them in forest planning also deserves further research.

Conclusions

Stand delineation with a cellular automaton produces stand boundaries that closely follow the borders between mineral soil and peat land. The degree of explained variation is also quite high for fertility class and growing stock variables, but a part of the stands might be mixtures of two or more fertility classes. Stand delineation created by cellular automaton is not sensitive to the set of stand variables that are used in the delineation. However, use of stand basal area and mean diameter is recommendable, together with soil type and fertility class. Optimization of the parameters of cellular automaton is not necessary in the practical use of the method. Virtual stands, suggested in the current study, provide a means to reduce biases in results calculated for heterogeneous segments.

Abbreviations

ALS: Airborne laser scanning; CA: Cellular automaton

Acknowledgments

Not applicable.

Authors' contributions

TP conducted the analyses and wrote the manuscript. The author read and approved the final manuscript.

Funding

No external funding received.

Availability of data and materials

The analysis is based on data publicly available at http://www.metsaan.fi/paikkatietoaineistot.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The author declares that he has no competing interests.

Received: 6 September 2019 Accepted: 26 February 2020
Published online: 12 March 2020

References

Baatz M, Schäpe A (2000) Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: Strobl J, Griesebner G, Blaschke T (eds) Angewandte Geograpische Informationsverarbeitung XII. Beiträge zum AGIT symposium, Salzburg, 22-23 June. Karlsruhe: Herbert Wichmann

Bettinger P, Graetz D, Boston K, Sessions J, Chung W (2002) Eight heuristic planning techniques applied to three increasingly difficult wildlife planning problems. Silv Fenn 36(2):561–584

Borges J, Hoganson R (1999) Assessing the impact of management unit design and adjacency constraints on forestwide spatial conditions and timber revenues. Can J For Res 29:1764–1774

Heinonen T, Kurttila M, Pukkala T (2007) Possibilities to aggregate raster cells through spatial optimization in forest planning. Silv Fenn 41(1):89–103

Heinonen T, Pukkala T (2007) The use of cellular automaton approach in forest planning. Can J For Res 37:2188–2200

Hyyppä J, Hyyppä H, Leckie D, Gougeon F, Yu X, Maltamo M (2008) Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. Int J Remote Sens 29:1339–1366

Jumppanen J, Kurttila M, Pukkala T, Uuttera J (2003) Spatial harvest scheduling approach for areas involving multiple ownership. Forest Policy Econ 5:27–38

Kangas A, Mehtätalo L, Mäkinen A, Vanhatalo K (2011) Sensitivity of harvest decisions to errors in stand characteristics. Silva Fennica 45:693–709

Kansanen K, Vauhkonen J, Lähivaara T, Seppänen A, Maltamo M, Mehtätalo L (2019) Estimating forest stand density and structure using Bayesian individual tree detection, stochastic geometry, and distribution matching. ISPRS J Photogr Remote Sens 152:66–78

Kansanen K, Vauhkonen J, Lähivaata T, Mehtätalo L (2016) Stand density estimators based on individual tree detection and stochastic geometry. Can J For Res 46(11):1359–1366. https://doi.org/10.1139/cjfr-2016-0181

Koch B, Kattenborn T, Straub C, Vauhkonen J (2014) Segmentation of forest to tree objects. In: Maltamo M, Næsset E, Vauhkonen J (eds) Forestry applications of airborne laser scanning: concepts and case studies. Managing Forest ecosystems 27. Springer Science+Business Media B.V. Dordrecht, The Netherlands, pp 89–112. https://doi.org/10.1007/978-94-017-8663-8__4

Koch B, Straub C, Dees M, Wang Y, Weinacker H (2009) Airborne laser data for stand delineation and information extraction. Int J remote Sens 30(4):935–963
Lu F, Eriksson LO (2000) Formation of harvest units with genetic algorithms. For Ecol Manag 130:57–67
Mustonen J, Packalen P, Kangas A (2008) Automatic segmentation of forest stands using a canopy height model and aerial photography. Scand J For Res 23:534–545. https://doi.org/10.1080/02827580802552446
Næsset E (2014) Area-based inventory in Norway – from innovation to an operational reality. In: Maltamo M, Næsset E, Vauhkonen J (eds) Forestry applications of airborne laser scanning: concepts and case studies. Managing Forest ecosystems 27. Springer Science+Business Media B.V. Dordrecht, The Netherlands, pp 215–240. https://doi.org/10.1007/978-94-017-8663-8__11
Olofsson K, Holmgren J (2014) Forest stand delineation from lidar point-clouds using local maxima of the crown height model and region merging of the corresponding Voronoi cells. Remote Sens Lett 50(3):268–276
Packalén P, Heinonen T, Pukkala T, Vauhkonen J, Maltamo M (2011) Dynamic treatment units in Eucalyptus plantations. For Sci 57(5):416–426
Pascual A, Pukkala T, de Miguel S, Pesonen A, Packalen P (2018) Influence of size and shape of forest inventory units on the layout of harvest blocks in numerical forest planning. Eur J Forest Res. https://doi.org/10.1007/s10342-018-1157-5
Pukkala T (1990) A method for incorporating the within-stand variation into forest management planning. Scand J Forest Res 5:263–275
Pukkala T (2018) Instructions for optimal any-aged forestry. Forestry 91(5):563–574. https://doi.org/10.1093/forestry/cpy015
Pukkala T (2019a) Using ALS raster data in forest planning. J For Res 30(5):1581–1593. https://doi.org/10.1007/s11676-019-00937-6
Pukkala T (2019b) Optimized cellular automaton for stand delineation. J For Res 30(1):107–119
Pukkala T, Kolström T (1991) Effect of spatial pattern of trees on the growth of Norway spruce stand. A simulation model. Silv Fenn 25(3):117–131
Pukkala T, Lähde E, Laiho O (2013) Species interactions in the dynamics of even- and uneven-aged boreal forests. J Sust For 32:1–33
Pukkala T, Miina J (2005) Optimising the management of a heterogeneous stand. Silv Fenn 39(4):525–538
Pukkala T, Packalén P, Heinonen T (2014) Dynamic treatment units in forest management planning. Manag Forest Ecosyst 33:373–392
Roncat A, Morsdorf F, Briese C, Wagner W, Pfeifer N (2014) Laser pulse interaction with forest canopy: Geometric and radiometric issues. In: Maltamo M, Næsset E, Vauhkonen J (eds) Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies. Managing Forest Ecosystems 27. Springer Science+Business Media, B.V. Dordrecht, The Netherlands, pp 19–41. https://doi.org/10.1007/978-94-017-8663-8_2
Siipilehto J (1999) Improving the accuracy of predicted basal-area diameter distribution in advanced stands by determining stem number. Silv Fenn 33(4):281–301
Vauhkonen J, Maltamo M, McRoberts RE, Næsset E (2014) Introduction to forestry applications of airborne laser scanning. In: Maltamo M, Næsset E, Vauhkonen J (eds) forestry applications of airborne laser scanning: concepts and case studies. Managing Forest ecosystems 27. Springer science+business media, B.V. Dordrecht, the Netherlands, pp 1–16. https://doi.org/10.1007/978-94-017-8663-8_1
Vauhkonen J, Mehtätalo L (2015) Matching remotely sensed and field measured tree size distributions. Can J For Res 45(3):353–363. https://doi.org/10.1139/cjfr-2014-0285
Vauhkonen J, Tokola T, Packalen P, Maltamo M (2009) Identification of single-tree attributes using airborne laser scanning-based height, intensity and alpha shape metrics. For Sci 55:37–47
Von Neumann J, Burks AW (1966) Theory of self-reproducing automata. Urbana: University of Illinois Press, Urbana and London, p 388
Wing BM, Boston K, Ritchie M (2018) A technique for implementing group selection treatments with multiple objectives using an airborne lidar-derived stem map in a heuristic environment. For Sci 65(2):211–222
Wolfram S (2002) A new kind of science. Wolfram Media, Champaign, Illinois. ISBN 1-57955-008-8. P 1280
Wu Z, Heikkinen V, Hauta-Kasari M, Parkkinen J, Tokola T (2013) Forest stand delineation using a hybrid segmentation approach based on airborne laser scanning data. In: Kämäräinen JK, Koskela M (eds) Image analysis. SCIA 2013. Lecture notes in computer science, vol. 7944. ISBN 978-3—642-38885-9. Springer, Berlin, Heidelberg, pp 95-106
Wulder MA, White JC, Hay GJ, Castilla G (2008) Towards automated segmentation of forest inventory polygons of high spatial resolution satellite imagery. Forest Chron 84(2):221–230