forests

MDPI

Article

Improved Cellular Automaton for Stand Delineation

Weiwei Jia [1], Yusen Sun [1], Timo Pukkala [1,2,*] and Xingji Jin [1]

1 Key Laboratory of Sustainable Forest Ecosystem Management-Ministry of Education, School of Forestry, Northeast Forestry University, Harbin 150040, China; jiaww2002@163.com (W.J.); sunyusen@nefu.edu.cn (Y.S.); xingjijin@163.com (X.J.)

2 School of Forest Sciences, Joensuu Campus, Faculty of Science and Forestry, University of Eastern Finland, P.O. Box 111, 80101 Joensuu, Finland

* Correspondence: timo.pukkala@uef.fi; Tel.: +358-40-511-2089

Received: 22 November 2019; Accepted: 21 December 2019; Published: 25 December 2019

Abstract: Airborne laser scanning (ALS) is becoming common in forest inventories. The data obtained by laser scanning contain the locations of the echoes of laser pulses. If these data are used in forest management, they need to be segmented into spatially continuous stands that are homogeneous in terms of stand attributes. Prior to segmentation, the laser pulse data can be processed into canopy height model, which shows the distance of canopy surface from the ground. This study used a cellular automaton with a canopy height model for the delineation of tree stands, considering three criteria: homogeneity of the stand in terms of growing stock attributes, stand area, and stand shape. A new method to consider stand shape in cellular automaton was presented. This method had a clear beneficial effect on the stand delineation result. Increasing weight of the shape criterion led to more roundish and less irregular stand shapes. Also, increasing weight of the stand area improved the shape of the stands. The cellular automaton led to average stand areas of 1–1.7 ha, depending on cell size and the parameters of the automaton. The cellular automaton explained 84.7–94.2% of the variation in maximum canopy height when 5 m × 5 m cells were used. Cell sizes of 5–10 m were found to result in the best stand delineation results.

Keywords: self-organizing system; forest map; grid data; laser scanning; ALS inventory; segmentation

1. Introduction

The use of airborne laser scanning (ALS) is increasing in forest inventories [1]. ALS can be conducted from manned aircrafts or unmanned aerial vehicles. The use of these data in traditional forest management requires a segmentation step where the data are organized into spatially continuous areas that correspond to tree stands.

Tree stands are usually understood as homogeneous subareas of the forest. Tree height, stand density, tree species composition, site fertility, etc. should be more or less the same within a stand. This makes it possible to have "permanent" stands, as treatments and future development are similar for the whole stand.

The traditional way of delineating tree stands is visual (and manual) where stand borders are drawn on aerial photographs, using perhaps soil maps, canopy height models, etc. as additional information. Visual delineation is gradually being replaced by automated computer algorithms, which may use the same data sources as used in traditional delineation [2]. Many methods were developed for numerical stand delineation [3–8]. These methods are often called segmentation methods, since the created areas do not necessarily correspond to traditional stand compartments. They might be too small for the implementation of forest management actions, which means that the segments must be further aggregated to obtain large enough continuous treatment blocks [9,10].

Many of the segmentation techniques developed for forestry are case- or task-specific [5–8]. As an alternative to task-specific methods, existing multi-purpose algorithms such as cellular automata and metaheuristics could also be used. The theoretical foundations of these methods might be stronger than that of case-specific methods, and there is much research from several fields on the multi-purpose algorithms.

Cellular automata (CA) are examples of multipurpose algorithms [11,12]. In forestry, CA were used to solve spatial forest planning problems [13–15], in land-use planning [16,17], and to simulate the spread of forest pests [18,19] and diseases [20]. Recently, they were also successfully used for stand delineation [8,10]. CA are self-organizing systems where the "state" of a cell depends on its neighborhood. The effect of other cells on a cell's state decreases rapidly with distance [16]. When CA are used for stand delineation, "cell state" corresponds to stand identification (ID) number. The purpose is to find such a stand number for each cell that cells belonging to the same stand form large enough continuous areas and their shape is "attractive". The cells that constitute a stand should be similar in terms of stand and site variables.

Commercial companies often do the laser scanning, and they may also preprocess the data. For example, the company may use the scanning data to develop a canopy height model (CHM), i.e., the difference between canopy surface and the ground. The CHM often indicates the canopy height on a grid of small cells, for instance, in 1-m resolution.

The canopy height model may be the only ALS product that the forest manager uses. The CHM is a valuable aid in visual stand demarcation or in the assessment of tree height. However, the CHM contains little numerical information directly applicable to forest management planning. For example, canopy height is not equal to tree height since a single tree crown often covers many cells of the CHM. In a mature stand of 400 trees per ha, a tree occupies on average 25 cells. In such a forest, the maximum value of the CHM in a 5×5 window might be used as an estimate of the tree height. Even that value can underestimate tree height since the laser pulses do not hit all treetops.

Due to the fact that the values of very small cells of a CHM do not measure stand height (not even tree height), there may be no reason to use the canopy height model directly for stand delineation. Instead, the values of the cells can be used to derive other variables like the maximum height within, for instance, a 5 m $\times$ 5 m window, which is an estimate of local tree height. The difference between maximum and minimum height is a measure of canopy depth. The difference between maximum and mean height, or between mean and minimum height, conveys information about crown shape and stand density (Figure 1). Therefore, using the maximum, mean, and minimum height of the CHM within a certain window takes into account several features of the stand, and it may, therefore, lead to better stand delineation than using only one variable (e.g., the mean canopy height in a cell).

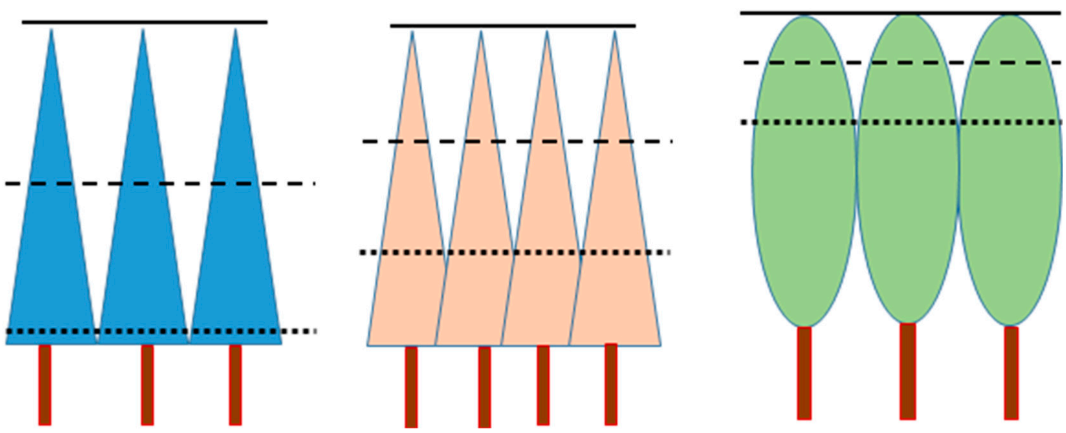


Figure 1. Three different canopies with the same maximum height (continuous horizontal lines) but different mean (dashed lines) and minimum (dotted lines) heights. Combined use of maximum, mean, and minimum heights helps to detect differences in stand density and crown form.

Replacing the values of the CHM by the mean, maximum, or minimum values of a window makes it possible to use a larger cell size, compared to the original CHM. Large cells allow faster calculations because the number of cells is reduced. Variables calculated for larger cells may also correlate better with stand variables. However, the use of large cells brings the mixed cell problem. A large cell might extend over more than one stand when located at the stand boundary. The mean canopy height of the enlarged cells is the average of two dissimilar stands. The maximum height of a stand of tall trees extends to adjacent stands, as does the minimum height of stands of short trees (Figure 2).

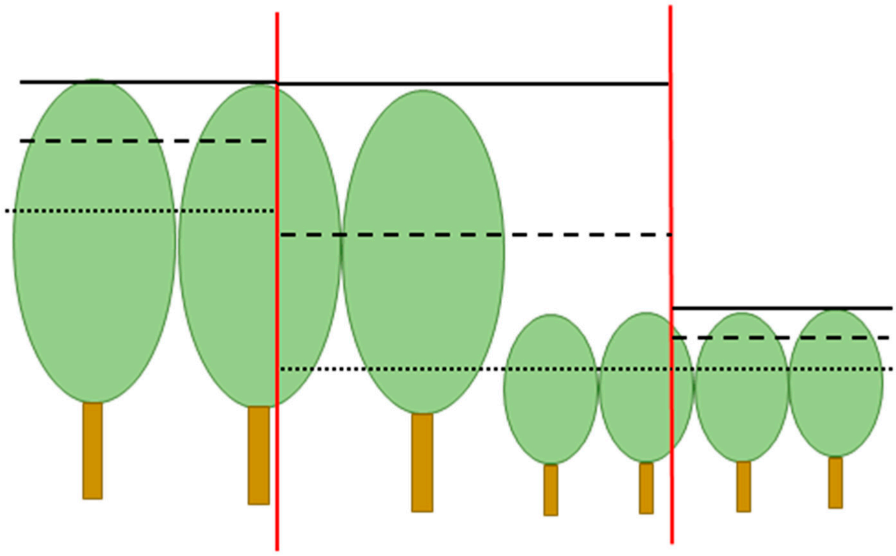


Figure 2. Maximum (continuous line), mean (dashed line), and minimum (dotted line) canopy height in a mixed cell consisting of two dissimilar stands (the area between the two vertical lines represents a mixed cell). In a mixed cell, the mean canopy height does not describe either of the stands; maximum height is taken from the taller stand, and minimum height is taken from the shorter stand.

There are typically three criteria in stand delineation: (1) stands should be homogeneous in terms of site and growing stock variables, (2) they should be large enough, and (3) stand shape should be acceptable (roundish, not too irregular). No overall measures and target levels exist for these criteria as the importance and the indicators of the criteria are subjective [21]. However, homogeneity means that within-stand variation in stand variables should be small compared to between-stand variation. This can be measured with the R^2 statistic, which is the proportion of total variation explained by the stand delineation [3,8].

For the area of the stand, "the bigger the better" might not be true [8]. A more common objective is to avoid small stands since they lead to too small and isolated harvest blocks and other treatment units. Another objective might be to have nearly equally sized stands and avoid very large stands. If stands are understood as indivisible treatment units, very large stands make it difficult to cut the same volume every year, for example. Regular stand shapes resembling squares, rectangles, circles, or hexagons are most probably favored by most forest managers.

The shape of the stand was not explicitly considered in previous research on the use of CA in stand delineation [8,10]. Earlier studies employed stand variables interpreted for grid cells as the basis of the delineation. In the current study, we investigate how well a canopy height model derived from ALS data is suited for automated stand delineation with self-organizing CA. The methods developed in earlier research are enhanced so that the shape of the stand is also explicitly included in the stand delineation process, in addition to stand area and homogeneity. The new variant of CA developed in this study is applied to a CHM of the Mengiagang forestry farm, located in the Heilogjiang province of China.

The objective of our study was to develop a CA for automated numerical stand delineation considering the homogeneity, area, and shape of the stands. Using the new algorithm, we analyzed the effect of increasing importance of stand area and stand shape on the outcome of the CA, compared to the case where all weight was on within-stand homogeneity. In addition, we analyzed the effect of enlarging the cell size on the delineation result.

Different stand delineations were compared in terms of R^2. Minimum, mean, and maximum stand area, proportion of small stands, and two form indices were calculated for the delineations. They were also assessed visually with the help of stand maps. It was assumed that one hectare is a sufficient stand size, and smaller stand sizes were penalized. Circular stand shape was assumed to be ideal. Small within-stand variation and large between-stand variation were pursued.

2. Materials and Methods

2.1. Case Study Forest

The study area was the Mengjiagang forest farm (46°32′ north (N), 129°10′ east (E)), located in Huanan County, Heilongjiang Province, China (Figure 3). The major planted tree species in this farm include *Pinus koraiensis* (henceforth referred to as Korean pine), *Picea asperata*, *Pinus sylvestris* (Mongolian pine), and *Larix gmelinii* (larch). The total area of the forest is 15,503 hectares, of which 4438 hectares are natural forests, accounting for 32.8% of the forest area.

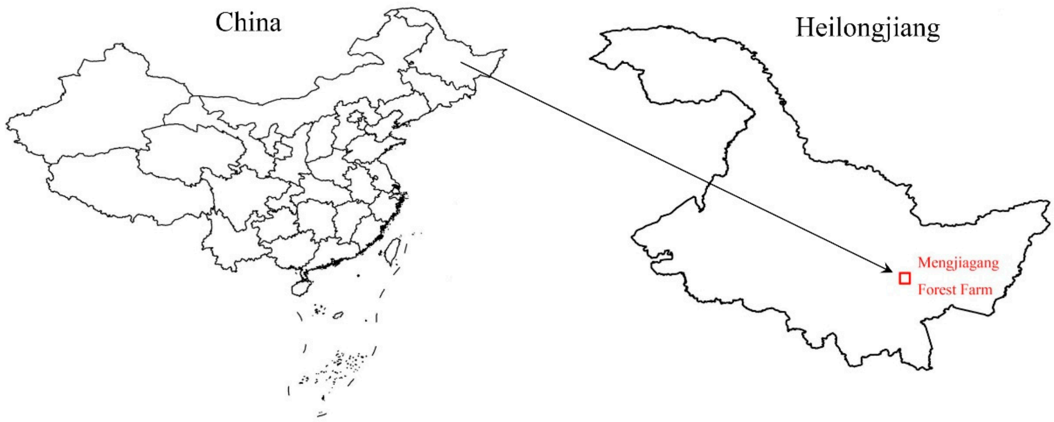


Figure 3. Location of the Mengjiagang forest farm in the Heilongjiang province of northeastern China.

2.2. Canopy Height Model

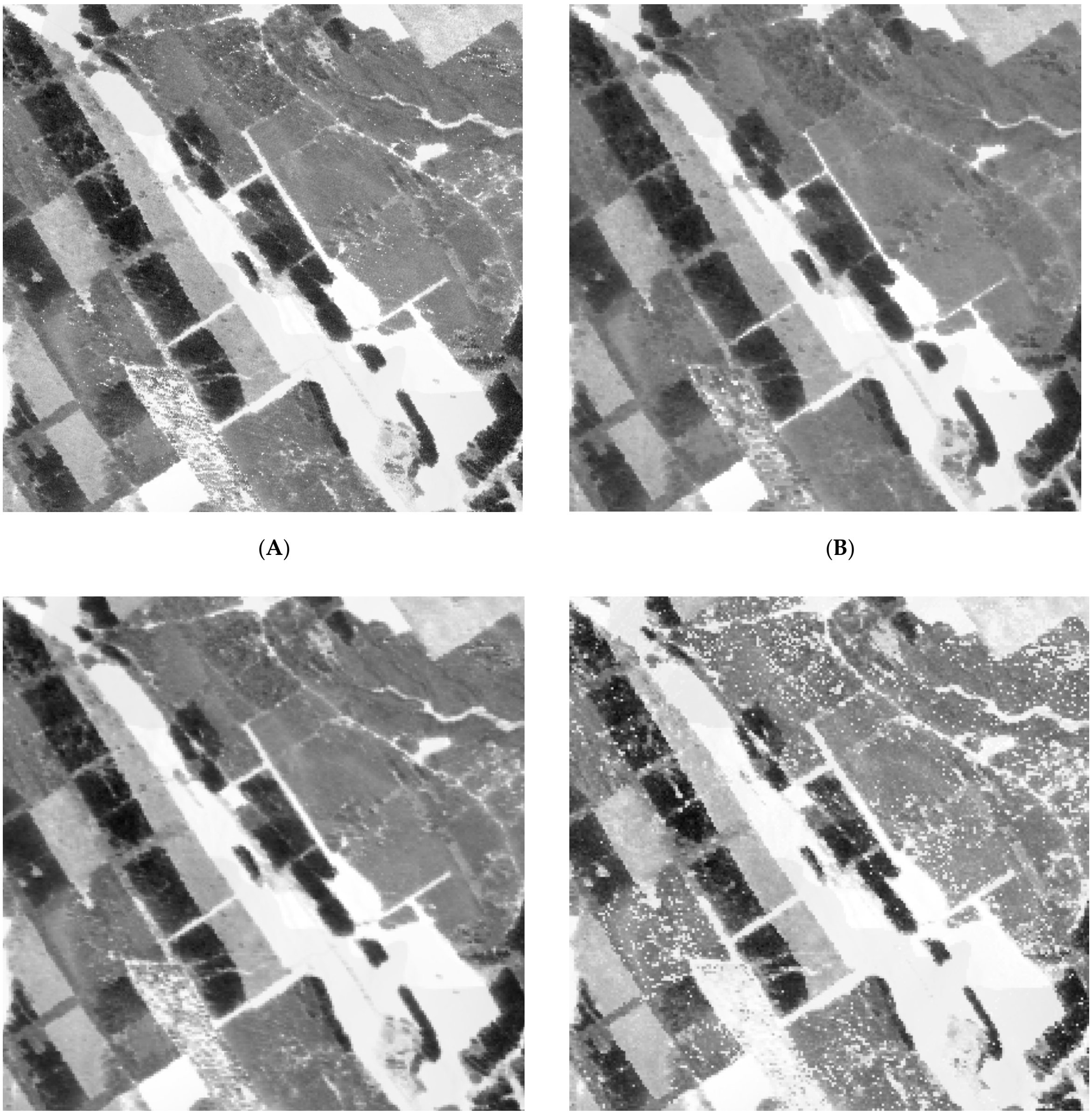
The airborne laser scanning was conducted between 31 May and 15 June in 2017. The flight altitude was 1000 m, with a pulse frequency of 300 kHz and a scanning angle of 30° in both directions from vertical. The average number of echoes per square meter was about four. The Institute of Forest Resource Information Techniques of the Chinese Academy of Forestry conducted the laser scanning. It also developed the canopy height model in October 2017. A digital elevation model (DEM) was generated from the ground echoes, and a digital surface model (DSM) was generated from the canopy echoes. The canopy height model (CHM) was obtained as the difference between the DSM and DEM [22]. The resolution of the canopy height model was 1 m.

2.3. Data Preprocessing

A 2 km × 2 km subarea (2000 by 2000 cells) of the CHM of the Mengjiagang forest farm was selected for the analyses of this study. The main criteria for choosing the sub-area were that most of it had to be forest, there was variation in tree species and canopy height, and the area included both instant and gradual changes in canopy height (clear and unclear stand borders).

As the first step of data processing, the CHM was checked for outliers. A few cells were found where canopy height was negative, and these values were replaced by the mean canopy height of a window of 5 × 5 cells. There were no illogically high values. The maximum canopy height was 31.17 m.

Then, every third, fifth, seventh, etc. row and column were selected (until 13), and the maximum, mean, and minimum canopy heights were calculated from a window of 3 × 3, 5 × 5, 7 × 7, etc. cells (Figure 4). When the window size (and the cell size of the output layer) was 3 × 3, the first row and column number was two, with the 5 × 5 window, it was three, etc.



(**A**) (**B**) (**C**) (**D**)

Figure 4. A 500 m × 500 m subarea of the case study forest. (**A**) is the original canopy height model (CHM) (cell size 1 m); (**B**) shows the maximum, (**C**) shows the mean, and (**D**) shows the minimum canopy height in 5 m × 5 m cells. Dark tone implies high canopy height.

A species raster was created from an existing forest map [23], where different species of the forest were coded as follows: 1 = Korean pine, 2 = Mongolian pine, 3 = larch, 4 = other species. The species layer was split into four 0–1 layers, the first indicating whether (1) or not (0) the species was Korean pine, the second indicating Mongolian pine, the third larch, and the fourth other species. If the values of all species layers were zero, the cell was not forest. The splitting was done because the species codes do not represent an interval scale, making it impossible to use Euclidean distance as a numerical measure of the similarity of tree species.

As a result of preprocessing, seven variables were available for each cell: maximum, mean, and minimum canopy height, as well as the presence of Korean pine, Mongolian pine, larch, or another tree species. These seven growing stock variables were used in stand delineation. All layers were standardized to mean zero and standard deviation one to remove the effect of different units (mean was first subtracted from the original value, and the result was divided by standard deviation).

2.4. Cellular Automaton

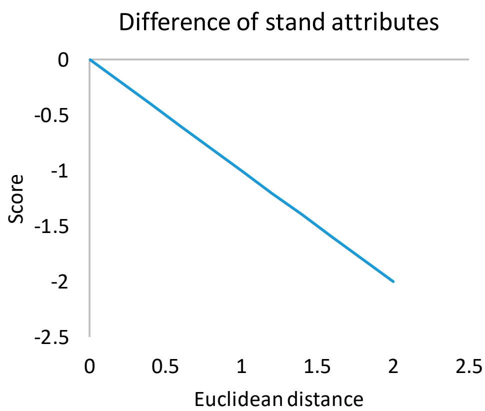
The purpose of the cellular automaton was to find the optimal or best stand number (ID number of the stand) for each cell of a grid. All cells were evaluated systematically, and the most suitable stand number was given to the cell. A cell always took the stand number of one of its adjacent cells. The number of adjacent cells was eight, since corner cells to the northeast, southeast, southwest, and northwest were also considered adjacent.

The most suitable stand number was selected for every cell of the grid for several iterations, until the stand borders no longer changed (or changed only little). The process was started from an initial stand delineation, which in this study consisted of 1-ha square-shaped stands. All cells that were within the 1-ha square were given the same stand number. A cell that has no adjacent cells with the same stand number disappears during the process since the cell gets the stand number from one of its adjacent cells. The stands may become divided into two or more disconnected parts during the CA run. More detailed descriptions of the use of CA for stand delineation can be found from Pukkala [8,10].

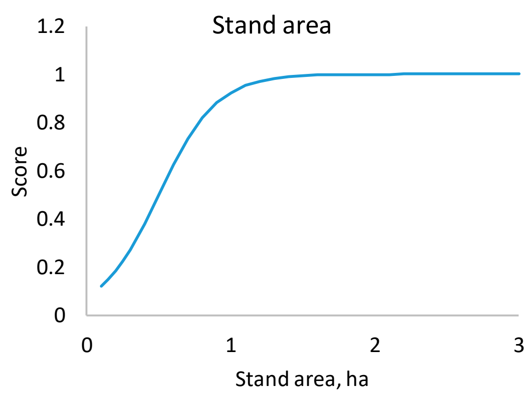
The function that was used to select the stand number for a cell was as follows:

$$P_{ij} = v_1 p_1\left(D_{ij}\right) + v_2 p_2\left(A_j\right) + v_3 p_3\left(B_{ij}\right) + v_4 p_4\left(S_{ij}\right), \quad (1)$$

where P_{ij} is the priority, or score, if cell i is joined to stand j, D_{ij} is the Euclidean distance of stand attributes between cell i and stand j, A_j is the area of stand j, B_{ij} is the proportion of common border between cell i and stand j (of the total border length of cell i), S_{ij} is the effect of joining cell i to stand j on the shape of stand j, p_k is sub-priority function for criterion k (Figure 5), and v_k is the weight of criterion k. The sum of the weights was equal to one. The score was calculated for each stand adjacent to cell i, and the number of the stand having the highest score was given to cell i. When calculating the border length, we assumed that the side borders of a cell (border with cells to east, west, south, and north) have a length equal to one and corner cells have a length equal to 0.3.

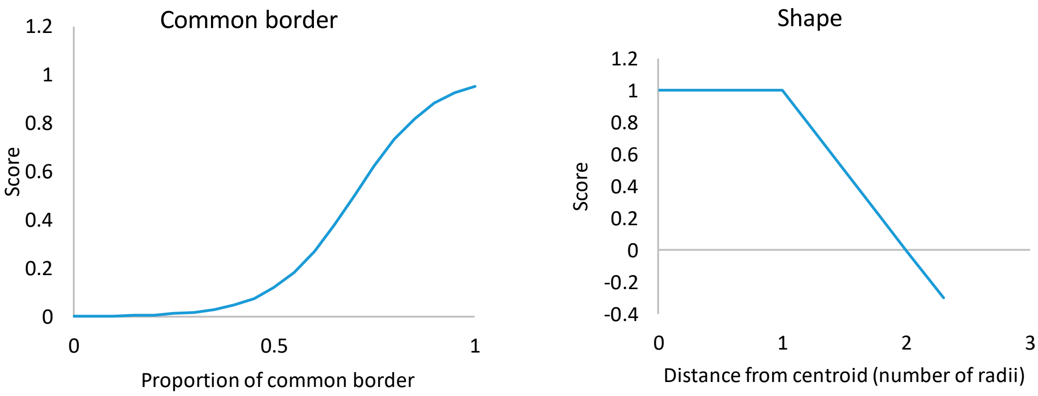


(A)



(B)

Figure 5. *Cont.*



(**C**) (**D**)

Figure 5. Sub-priority functions for the Euclidean distance of stand attributes between a cell and a stand (**A**), stand area (**B**), proportion of common border between a cell and a stand (**C**), and distance of cell from stand centroid (**D**). Distance from centroid is expressed as the number of radii of a circle that has the same surface area as the stand. The diagrams show the shapes of sub-priority functions p_1, p_2, p_3, and p_4 of Equation (1).

The idea of calculating the shape measure S_{ij} was that cells within a radius of a circle centred on the centroid of stand j and having the same area as stand j are not penalized because these cells do not deteriorate the stand's shape, compared to the ideal circular shape (Figure 6). If a cell is further from the centroid than the radius of a circular shape, the stand is penalized, and the further the cell is from the stand centroid.

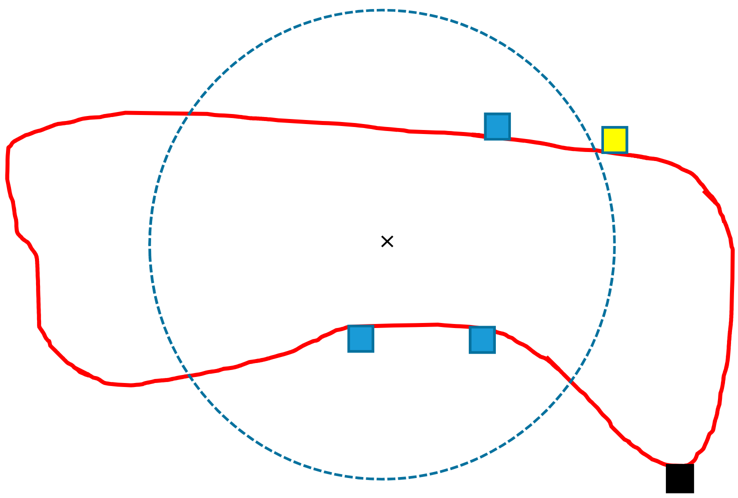


Figure 6. The principle of taking stand shape into account in the cellular automaton. Symbol × is the centroid of the stand shown with a red line. The dotted line is a circle with the same center and area as the stand. Small rectangles are cells of a grid. Joining the blue cells to the stand produces full shape points since they are within the radius of the circle. The yellow cell gets a small shape penalty, and the black cell gets a large penalty because these cells have an adverse effect on the shape of the stand.

The proportion of common border between cell i and stand j (B_{ij}), as well as the area of stand j (A_j), had a nonlinear effect on the score (Figure 5). The sigmoid type of relationship between stand area and score led to the avoidance of small stands, but stand area ceased to increase the score after about 1.5 ha. This means that very large stands were not considered better than 1.5-ha stands. A sigmoid relationship between common border (B_{ij}) and score led to avoiding stands that were composed of narrow strips of cells [8]. The common border criterion affected stand shape at the local cell level, whereas the stand shape criterion took into account the overall shape of the whole stand.

The sigmoid-type relationships were described with logistic functions.

$$p_2 = \frac{1}{\exp\left\{a_1\left(A_j - a_2\right)\right\}}. \tag{2}$$

$$p_3 = \frac{1}{\exp\left\{b_1\left(B_{ij} - b_2\right)\right\}}. \tag{3}$$

The values of the parameters were $a_1 = -5$, $a_2 = 0.5$, $b_1 = -10$, and $b_2 = 0.7$. They resulted in the relationships shown in Figure 5. The distance of stand attributes between cell i and stand j was a weighted Euclidean distance.

$$D_{ij} = \sqrt{\sum_{k=1}^{7} w_k\left(x_{ik} - x_{jk}\right)^2}, \tag{4}$$

where x_{ik} is the value of standardized stand attribute k in cell i, x_{jk} is the average value of the same attribute in stand j, and w_k is the weight of attribute k. Index k refers to the used stand attributes as follows (the weight used in this study is in parentheses): 1 = maximum canopy height (0.4); 2 = mean canopy height (0.3); 3 = minimum canopy height (0.2); 4 = Korean pine (0.025); 5 = Mongolian pine (0.025); 6 = larch (0.025); 7 = other species (0.025).

2.5. Stand Delineation Cases

This article concentrated on analyzing the effect of stand shape (S_{ij}) and stand area (A_j) criteria on the stand delineation result (Table 1). Firstly, a reference delineation was produced where the only criterion was the similarity of stand variables within a stand (in Equation (1), v_1 was 1, and all the other weights were zero).

Table 1. Optimization cases. Dist.—distance; Max—maximum; Min—minimum; CH—canopy height.

Name	Cell Size m	v_1 Dist.	v_2 Area	v_3 Border	v_4 Shape	w_1 Max CH	w_2 Mean CH	w_3 Min CH	$\Sigma w_{4–7}$ Species
Reference	5	1	0	0	0	0.4	0.3	0.2	0.1
Shape 0	5	0.6	0.2	0.2	0	0.4	0.3	0.2	0.1
Shape 1	5	0.5	0.2	0.2	0.1	0.4	0.3	0.2	0.1
Shape 2	5	0.4	0.2	0.2	0.2	0.4	0.3	0.2	0.1
Shape 3	5	0.3	0.2	0.2	0.3	0.4	0.3	0.2	0.1
Area 0	5	0.6	0	0.2	0.2	0.4	0.3	0.2	0.1
Area 1	5	0.5	0.1	0.2	0.2	0.4	0.3	0.2	0.1
Area 2	5	0.4	0.2	0.2	0.2	0.4	0.3	0.2	0.1
Area 3	5	0.3	0.3	0.2	0.2	0.4	0.3	0.2	0.1
Cell 1	1	0.5	0.15	0.2	0.15	0.4	0.3	0.2	0.1
Cell 3	3	0.5	0.15	0.2	0.15	0.4	0.3	0.2	0.1
Cell 5	5	0.5	0.15	0.2	0.15	0.4	0.3	0.2	0.1
Cell 7	7	0.5	0.15	0.2	0.15	0.4	0.3	0.2	0.1
Cell 9	9	0.5	0.15	0.2	0.15	0.4	0.3	0.2	0.1
Cell 11	11	0.5	0.15	0.2	0.15	0.4	0.3	0.2	0.1
Cell 13	13	0.5	0.15	0.2	0.15	0.4	0.3	0.2	0.1

Then, the effects of stand shape (S_{ij} in Equation (1)) and stand area (A_j) criteria were analyzed by varying their weights. When the weight of stand area or shape was increased, the weight of the similarity of stand variables (D_{ij}) was decreased by the same amount such that the sum of criterion weights was always 1. Table 1 is a summary of the analyzed CA runs.

The last part of the analyses examined the effect of the size of the window from which the maximum, mean, and minimum canopy heights were calculated (cell size was equal to window size). The species layer was resampled to the same cell size by using the mode filter (the most common tree species within the window was selected). The analyzed cell sizes were 1 × 1, 3 × 3, 5 × 5, 7 × 7, 9 × 9, 11 × 11, and 13 × 13 square meters (the cell size of the original CHM was 1 × 1 m). The number of iterations was 20 in all CA runs since it was found that the delineation almost stabilized during 20 iterations.

The results of different CA runs were compared in terms of the degree of explained variance of the stand attributes (R^2). R^2 was calculated as

$$R^2 = 1 - \mathrm{SSE}/\mathrm{SST}, \quad (5)$$

where SSE is the variation not explained by the delineation, and SST is the total variation of the attribute within the grid. SST and SSE were calculated as follows:

$$\mathrm{SST} = \sum_{j=1}^{N} \sum_{i=1}^{n_j} \left(y_{ij} - \overline{y}\right)^2, \quad (6)$$

$$\mathrm{SSE} = \sum_{j=1}^{N} \sum_{i=1}^{n_j} \left(y_{ij} - \overline{y_j}\right)^2, \quad (7)$$

where N is the number of stands, n_j is the number of cells in stand j, y_{ij} is the value of the variable in cell i belonging to stand j, $\overline{y}$ is the overall mean of the variable, and $\overline{y}_j$ is the mean value of the variable among the cells that belonged to stand j.

In addition, the proportion of stands smaller than 0.1 ha and two variables that described the form of the stands were calculated for every delineation.

$$\text{Form 1} = \text{Stand perimeter (m)/Square perimeter (m)}. \quad (8)$$

$$\text{Form 2} = \text{Stand perimeter (m)}/\sqrt{(\text{Number of cells in the stand})}. \quad (9)$$

Stand perimeter is the total length of the outer border of the stand without any smoothing, and square perimeter is the perimeter of a square having the same area as the stand. The latter form index was called form heterogeneity in Baatz and Schäpe [21]. A small value of both indices implies “good” stand shape. The form indicators were calculated as unweighted averages of the stand values.

3. Results

3.1. *Effect of Stand Shape Criterion*

The results for the reference delineation were good in terms of R^2 (Table 2). High R^2 means that most variation in stand attributes was explained by the delineation. On the other hand, the reference CA run produced hardly acceptable stand shapes (Figure 7A).

Table 2. Results for initial stand delineation (1-ha square-shaped stands), reference delineation (all weight was on the minimization of within-stand variation), and three delineations where the weight of the stand shape criterion (v_4 in Equation (1) was gradually increased. “R^2 all” is the average R^2 of all seven stand attributes used in delineation (CH = canopy height).

Variable	Initial	Reference	Shape 0	Shape 1	Shape 2	Shape 3
		$v_4 = 0$	$v_4 = 0$	$v_4 = 0.1$	$v_4 = 0.2$	$v_4 = 0.3$
R^2, maximum CH	0.530	0.942	0.911	0.892	0.873	0.847
R^2, mean CH	0.520	0.946	0.914	0.891	0.871	0.841
R^2, minimum CH	0.408	0.827	0.715	0.690	0.671	0.645
R^2, all	0.580	0.875	0.845	0.827	0.809	0.786
Number of stands	400	394	280	289	295	287
Minimum area, ha	1.000	0.005	0.001	0.005	0.003	0.001
Mean area, ha	1.00	1.02	1.43	1.38	1.36	1.39
Maximum area, ha	1.00	9.6	5.58	4.52	4.27	4.68
Small stands, %	0.0	1.8	2.9	4.8	3.7	2.4
Form 1	107	598	266	206	175	157
Form 2	19	106	47	37	31	28

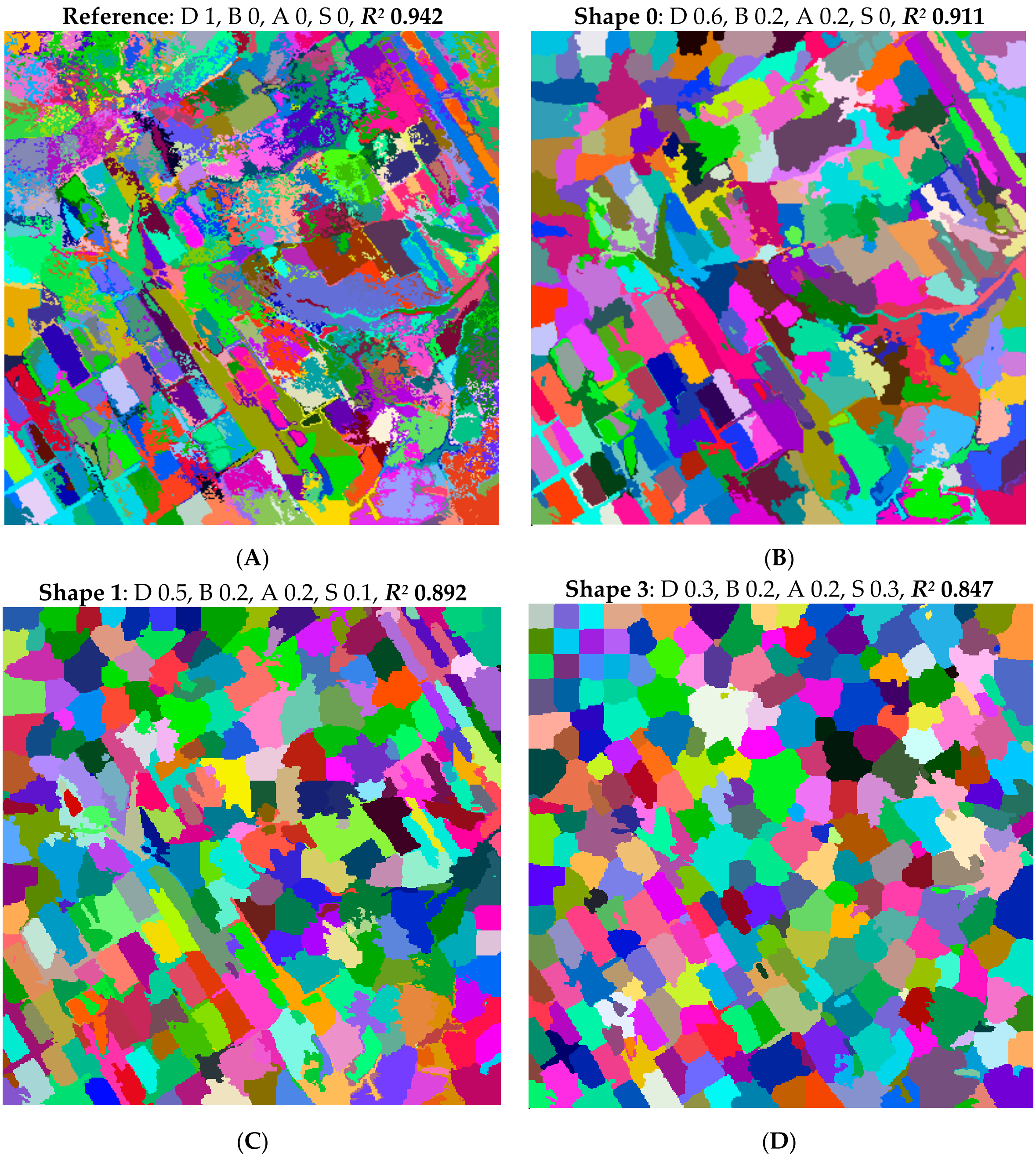


Figure 7. Effect of the stand shape criterion on the result of cellular automaton (CA) run (different colors indicate different stand numbers). (**A**): all weight was on small within-stand variation in stand attributes. The other CA runs included also other criteria, and the weight of stand shape was 0 (**B**), 0.1 (**C**), or 0.3 (**D**). R^2 is the degree of explained variance for maximum canopy height. Letters D, A, B, and S refer to the criteria of Equation (1) (D = distance, B = common border, A = area, S = shape), and the numbers indicate criterion weights.

Including more criteria in the function that was used to assign stand numbers to cells (Equation (1)) decreased R^2 but improved stand shape (Table 2, Figure 7B). The number of stands decreased by about 25%, and the average stand area increased by 40% (Shape 0 in Table 2). The maximum stand area decreased, which means that variation in stand size also decreased.

Giving more importance to the stand shape criterion had a clear effect on stand shape (Figure 7). A higher weight of the shape criterion produced more roundish stands. Increasing weight of the stand shape criterion decreased the number of narrow stands and the occurrence of narrow stand parts.

3.2. Effect of Stand Area Criterion

Also, increased weight of stand area (v_2 in Equation (1)) improved stand shape (Table 3, Figure 8). Good stand shape and within-stand homogeneity were a trade-off since the degree of explained variance decreased when the shape of the stands improved. Therefore, no optimal criteria weights can be given, as the weights depend on the preferences of the forest managers who use the delineation.

Table 3. Results for cellular automaton (CA) runs where the weight of stand area criterion (v_2 in Equation (1)) was gradually increased. "R^2 all" is the average R^2 of all seven stand attributes used in delineation (CH = canopy height).

Variable	Area 0	Area 1	Area 2	Area 3
	$v_2 = 0$	$v_2 = 0.1$	$v_2 = 0.2$	$v_2 = 0.3$
R^2, maximum CH	0.897	0.886	0.873	0.847
R^2, mean CH	0.898	0.885	0.871	0.841
R^2, minimum CH	0.702	0.684	0.671	0.645
R^2, all	0.841	0.830	0.809	0.786
Number of stands	396	365	295	238
Minimum area, ha	0.010	0.005	0.003	0.060
Mean area, ha	1.01	1.10	1.36	1.68
Maximum area, ha	3.04	3.70	4.27	4.86
Small stands, %	2.0	4.9	3.7	0.4
Form 1	198	183	174	168
Form 2	35	32	31	30

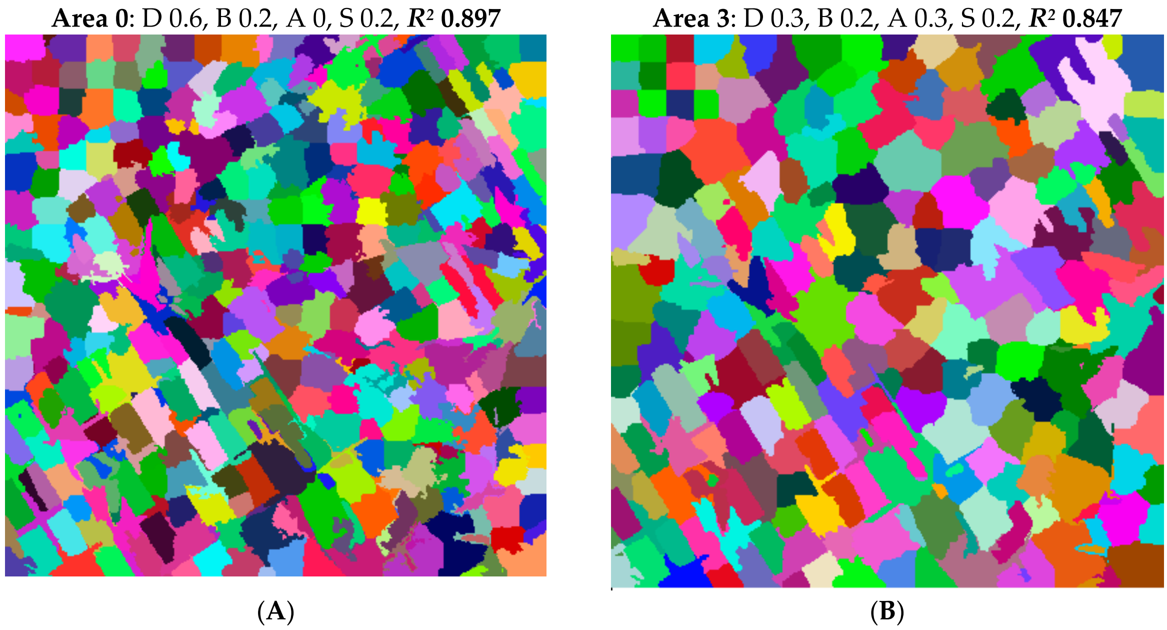


Figure 8. Effect of stand area criterion on the result of cellular automaton run (different colors indicate different stand numbers). The weight of stand area was 0 (**A**), or 0.3 (**B**). R^2 is the degree of explained variance for maximum canopy height. Letters D, A, B, and S refer to the criteria of Equation (1) (D = distance, B = common border, A = area, S = shape), and the numbers are criterion weights.

Compared to the case where the weight of stand area was zero ("Area 0" in Table 1), transferring weight from the similarity of stand attributes (D in Equation (1)) to stand area criterion (A) increased the average and maximum stand area, decreased the number of stands, and improved the values of stand form indices. The proportion of small (<0.1 ha) stands tended to decrease, but the trend was not systematic.

Visual inspection of the stand maps (Figure 9) revealed some differences that were not easily discernible from the numerical results of Table 3. In addition to improving stand shape, increasing weight on stand area led to a smaller number of small stands, and the stand borders became straighter. In particular, when the weight of the stand area criterion was zero, several stand borders were rugged, and some parts of the forest were divided into very small stands.

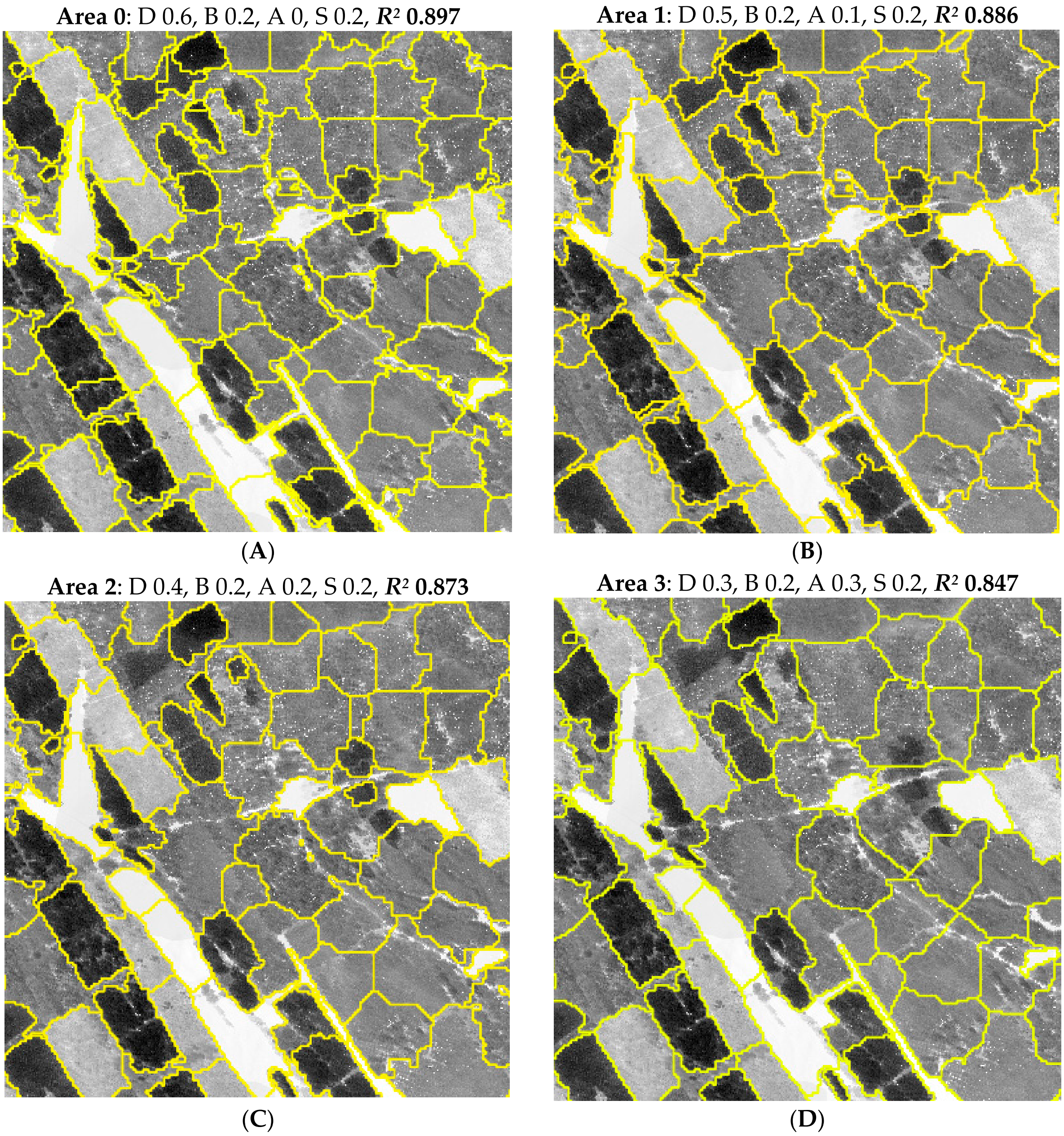


Figure 9. A part (500 m × 500 m) of the canopy height model (1-m resolution) with stand delineations obtained when the weight of stand area was 0 (**A**), 0.1 (**B**), 0.2 (**C**) or 0.3 (**D**).

3.3. *Effect of Cell Size*

Increasing cell size decreased the number of stands (Table 4). The minimum, mean, and maximum stand areas also increased, but the changes were not large and not always systematic. The proportion of small stands tended to decrease when cell size increased. Form indicator 1 (Equation (5)) indicated improving stand shape with increasing cell size. Form indicator 2 (Equation (6)) cannot be used to compare delineations obtained with different cell sizes.

Table 4. Effect of cell size on stand delineation. The number after "Cell" indicates the cell size in meters.

Variable	Cell 1	Cell 3	Cell 5	Cell 7	Cell 9	Cell 11	Cell 13
No. of stands	388	350	325	313	309	293	292
Minimum area, ha	1.013	0.010	0.005	0.003	0.020	0.060	0.030
Mean area, ha	1.03	1.14	1.23	1.27	1.29	1.35	1.35
Maximum area, ha	2.91	3.17	3.30	3.85	3.69	4.86	5.48
Small stands, %	3.4	3.4	5.8	2.6	2.9	0.3	0.7
Form 1	304	212	192	183	178	172	167

Visual inspection of the results revealed some additional effects of cell size (Figure 10). In particular, the use of the smallest 1-m cell size resulted in very long and rugged stand boundaries in some places. This problem was largely mitigated when the cell size was 5 m or more.

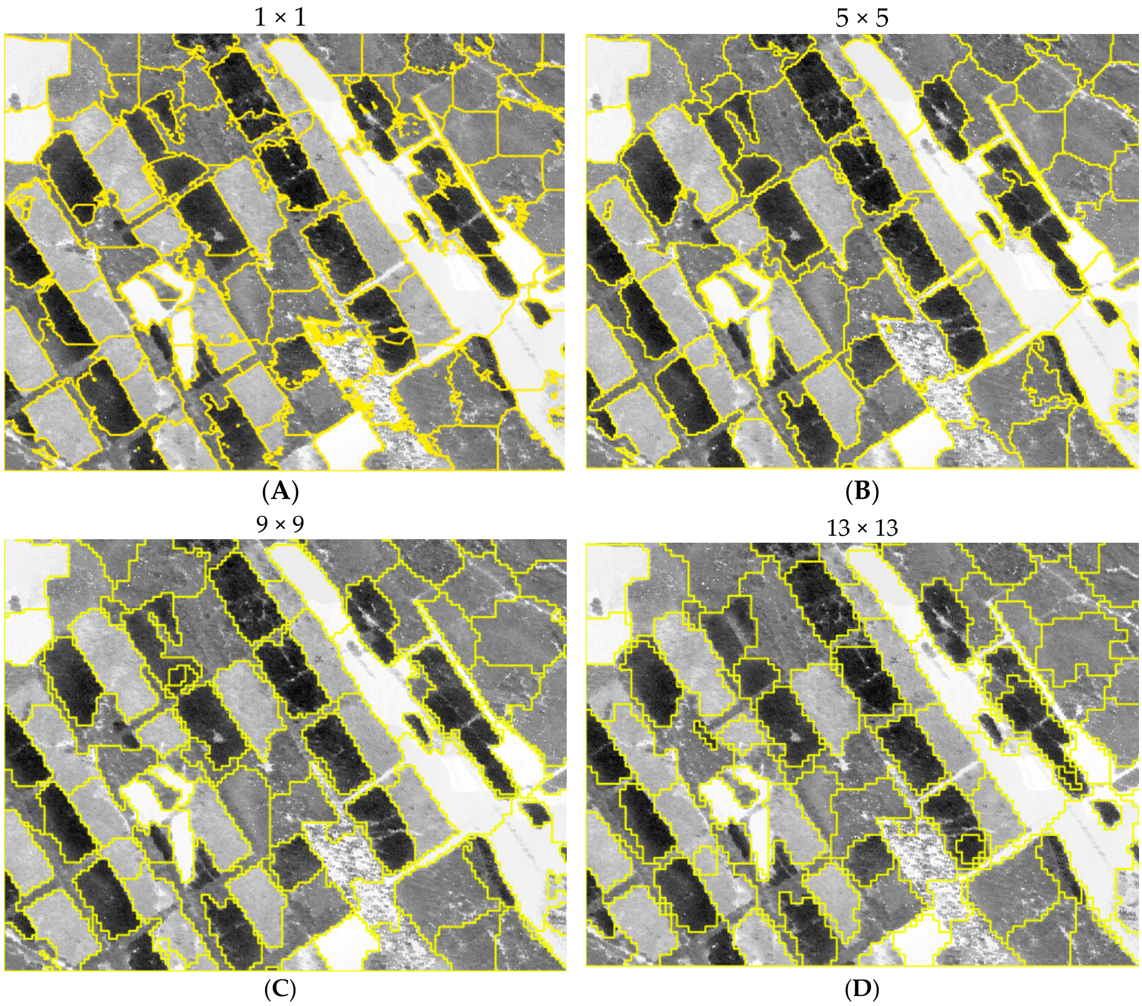
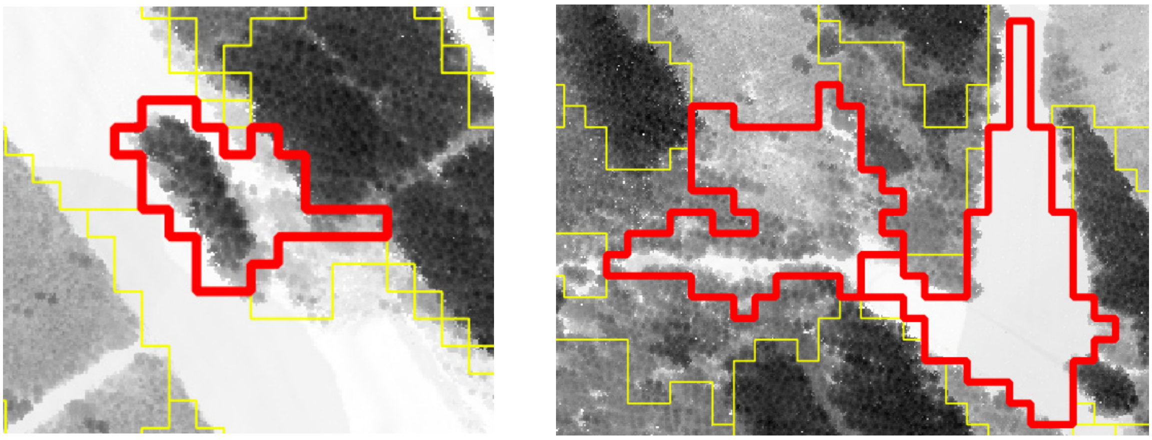


Figure 10. A part (about 500 m × 500 m) of the canopy height model (1-m resolution) with the stand delineations obtained when the cell size was 1 (**A**), 5 (**B**), 9 (**C**), or 13 m (**D**).

Comparison of the canopy height model and the stand borders showed that stands with tall trees were enlarged and some stands with short trees shrank, compared to the location of the stand edge in the canopy height model (Figure 10). This phenomenon can be seen more easily from Figure 11, which shows two details from the stand delineation obtained with 13-m cells. The overestimation of the area of stands with tall trees was mainly due to the high weight of maximum canopy height in Equation (4)

(Table 1) and the fact that the maximum canopy height of a mixed cell (cell located at stand border) is taken from that part of the cell where the trees are taller (see Figure 2).



(**A**) (**B**)

Figure 11. Details of stand delineation obtained with 13-m cell size. On the (**A**), a very small stand with much taller trees than in its surroundings is delineated too large. On the (**B**), the area of a stand with very low canopy height (light-gray tone) is underestimated, and narrow areas of low canopy height are connected to adjoining stands.

4. Discussion

Our study corroborated earlier findings that cellular automaton is an easy and flexible method capable of producing good stand delineation results [8,10]. Several criteria can be taken into account, depending on the preferences of the forest manager and the purpose of the delineation. As a difference to the previous studies, which used stand variables interpreted for rather large cells (16 m × 16 m), we used a high-resolution canopy height model developed from laser scanning data.

If there are several adjacent initial stands in a large homogeneous forest, the CA does not move the stand boundaries easily. This can be seen from Figure 7D, where there are several square-shaped stands in the northwestern corner. These squares are almost the same as the initial stands. This result is partly due to the shape of the sub-priority function for stand area, which assumed that 1-ha stand size was ideal, and additional area no longer increased the stand's area score. Linearizing the sub-priority function for stand area, using low weights for stand area and a common border, and letting the CA run longer would mitigate this problem.

Figure 12 shows that the CA fine-tuned for a certain area works equally well in areas other than the one which was used for fine-tuning. If required, the stand delineation produced by the CA can be finalized manually. Another option is to use the delineation as the starting point for another CA run, which uses different values for some parameters. For example, the somewhat rugged boundaries in some of the maps of Figure 12 could be rectified by changing the shape of the sub-priority function to common and increasing its weight and running the CA for a few iterations. Geographic information system (GIS) software packages also include tools for smoothing rugged stand borders.

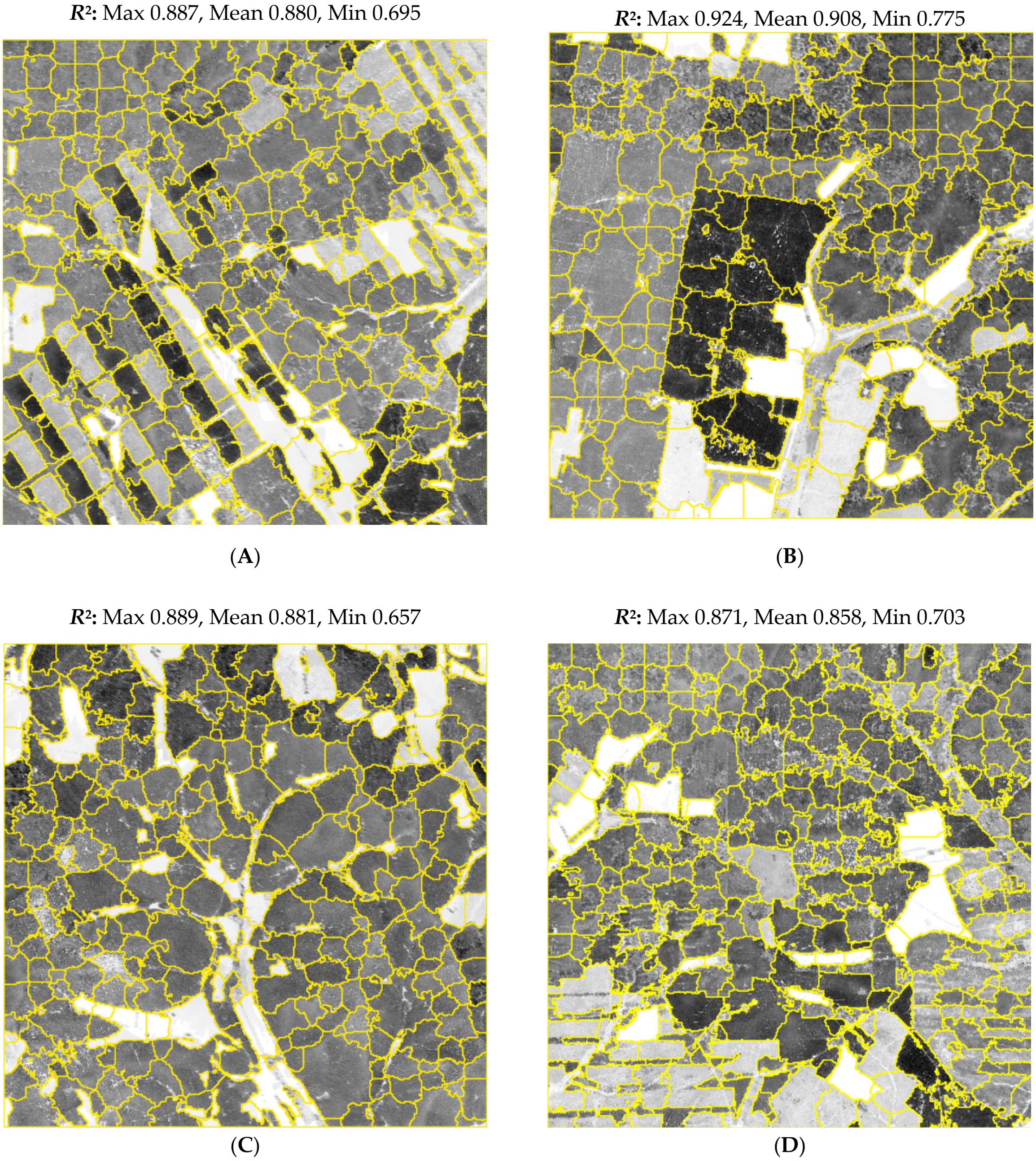


Figure 12. Delineations obtained for four areas (**A**–**D**) with the same CA parameters. The weights of the delineation criteria were v_1 = 0.4 (homogeneity), v_2 = 0.1 (area), v_3 = 0.2 (border), and v_4 = 0.3 (shape). The parameters of the sub-priority functions (Equations (2) and (3)) were: $a_1 = -1$, $a_2 = 1$, $b_1 = -3$, $b_2 = 0.5$. Species layers were not used in delineation. The R^2 statistics are shown for the maximum, mean, and minimum canopy height.

The spatial resolution of the CHM was 1 m. The small original cell size allowed us to compare different cell sizes in stand delineation. Since larger cells contained several initial cells, it was possible to calculate variables that described the range of variation in canopy height within short distance. We partitioned the single variable of the original canopy height model into three variables that conveyed more information about stand features than the canopy height alone. The variables used in this study were maximum, mean, and minimum canopy heights within the cell. It would be possible to calculate many more variables, such as standard deviation, skewness, and percentiles, but most probably the contribution of the additional variables to the delineation result would remain small.

It was assumed that large cell size leads to the mixed cell problem, where narrow artefact stands are delineated at stand borders [9]. However, the results showed that the mixed stand problem was not serious, most probably because the shape parameter prevented the creation of long and narrow stands. However, the other assumed consequence of large cell size, namely, the enlargement of stands of tall trees to adjacent stands (see Figure 2), was apparent when cell size was large (Figure 11). This result was partly due to the fact that the maximum canopy height had a larger weight in delineation than the mean and minimum height. High weight of minimum canopy height might have reversed the effect on enlarging cell size.

Tree species information was also used for stand delineation since it was assumed that tree species data are always available in plantation forests. However, the weight of the species layers was always low since we wanted to analyze the effect of other variables. The degree of explained variance for the four tree species layers was 0.8–0.9. Lack of perfect separation of stands with different species was most probably mainly due to the fact that the boundaries of the species map were straighter than the true stand borders, and they were not always exactly in the correct place.

The study developed a new way to consider stand shape on cellular automaton. The developed method worked as expected, and it had a clear improving influence on stand delineation when round-shaped stands with low perimeter/area ratio were targeted. Other shape indices for measuring stand shape were suggested in previous studies [21], such as border length divided by the square root of the number of cells (Equation (5)). However, some of the earlier indices measure the smoothness of the border more than stand shape. The shape criterion employed in our CA does not use border length at all but measures deviations from circular shape. In most cases, small irregularities in stand border can be smoothed out, which reduces the length of stand border but has no effect on the overall shape of the stand.

The results of our study showed that both stand area and stand shape should be used as criteria when CA are used for stand delineation since, otherwise, the delineation may not be acceptable. Small weights are sufficient since additional improvements in stand shape and area would result in increased within-stand variation in stand attributes.

Increasing cell size greatly reduces the number of cells. The use of large cells makes it possible to divide large canopy height models into stands in short computing time. The use of very small cells was found to be even harmful since small cells may result in very long and complicated stand borders. However, as discussed above, too large cells bring problems too. On the basis of our study, it may be concluded that the optimal cell size is 5–10 m when canopy height models are used for stand delineation.

The results of this study can be generalized to other regions since canopy height models derived from ALS data are not very different in different regions and countries. However, separating stands with different species compositions requires more attention when the stands are not planted monocultures. Combined use of maximum, mean, and minimum canopy height most probably helps to separate stands with different species compositions. Additional information can be obtained from other remote sensing material (for instance, aerial photographs). High-density laser scanning data can be used to calculate alpha shape metrics, which also help in species identification [24]. There are also other potential methods, based on statistical analysis and machine learning, which could be tested in species identification from laser data.

5. Conclusions

The study showed that cellular automata provide a flexible method capable of producing good stand delineation results. Several criteria can be taken into account, depending on the purpose of the delineation. The new method developed for considering stand shape in cellular automaton had a clear beneficial effect on the stand delineation result. Increasing weight of the shape criterion led to more roundish and less irregular stand shapes. Also, increasing weight of stand area improved the shape of the stands. The cellular automaton explained 84.7–94.2% of the variation of maximum canopy height when 5 m × 5 m cells were used. Cell sizes of 5–10 m were found to result in the best stand delineation results.

Author Contributions: Conceptualization, all authors; methodology, all authors; software, T.P.; formal analysis, Y.S.; resources, W.J.; data curation, W.J.; writing—original draft preparation, T.P.; writing—review and editing, all authors; visualization, Y.S.; project administration, W.J. and X.J.; funding acquisition, W.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China grant number 2017YFD0600404-2, the Fundamental Research Funds for the Central Universities grant number 2572019CP08, and the Heilongjiang Touyan Innovation Team Program (Technology Development Team for High-efficient Silviculture of Forest Resources).

Acknowledgments: We thank Pang Yong at the Chinese Academy of Forestry for sharing the canopy height data with us.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Vauhkonen, J.; Maltamo, M.; McRoberts, R.E.; Næsset, E.; Managing Forest Ecosystems. Introduction to forestry applications of airborne laser scanning. In *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies*; Maltamo, M., Naesset, E., Vauhkonen, J., Eds.; Springer: Dordrecht, The Netherlands, 2014; Volume 27, pp. 1–16.
2. Koch, B.; Kattenborn, T.; Straub, C.; Vauhkonen, J.; Managing Forest Ecosystems. Segmentation of forest to tree objects. In *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies*; Maltamo, M., Naesset, E., Vauhkonen, J., Eds.; Springer: Dordrecht, The Netherlands, 2014; Volume 27, pp. 89–112.
3. Mustonen, J.; Packalen, P.; Kangas, A. Automatic segmentation of forest stands using a canopy height model and aerial photography. *Scand. J. For. Res.* **2008**, *23*, 534–545. [CrossRef]
4. Wulder, M.A.; White, J.C.; Hay, G.J.; Castilla, G. Towards automated segmentation of forest inventory polygons of high spatial resolution satellite imagery. *For. Chron.* **2008**, *84*, 221–230. [CrossRef]
5. Koch, B.; Straub, C.; Dees, M.; Wang, Y.; Weinacker, H. Airborne laser data for stand delineation and information extraction. *Int. J. Remote Sens.* **2009**, *30*, 935–963. [CrossRef]
6. Wu, Z.; Heikkinen, V.; Hauta-Kasari, M.; Parkkinen, J.; Tokola, T. Forest stand delineation using a hybrid segmentation approach based on airborne laser scanning data. In *Image Analysis SCIA 2013 Lecture Notes in Computer Science*; Kämäräinen, J.K., Koskela, M., Eds.; Springer: Berlin, Germany, 2013; Volume 7944, pp. 95–106. ISBN 978-3-642-38885-9.
7. Dechesne, C.; Mallet, C.; Le Bris, A.; Gouet-Brunet, V. Semantic segmentation of forest stands of pure species combining airborne lidar data and very high resolution multispectral imagery. *ISPRS J. Photogramm. Remote Sens.* **2017**, *126*, 129–145. [CrossRef]
8. Pukkala, T. Optimized cellular automaton for stand delineation. *J. For. Res.* **2019**, *30*, 107–119. [CrossRef]
9. Pascual, A.; Pukkala, T.; de Miguel, S.; Pesonen, A.; Packalen, P. Influence of size and shape of forest inventory units on the layout of harvest blocks in numerical forest planning. *Eur. J. For. Res.* **2019**, *138*, 1–13. [CrossRef]
10. Pukkala, T. Using ALS raster data in forest planning. *J. For. Res.* **2019**, *30*, 1581–1593. [CrossRef]
11. Von Neumann, J. *Theory of Self-Reproducing Automata*; Burks, A.W., Ed.; University of Illinois Press: Urbana, IL, USA, 1966.
12. Wolfram, S. *A New Kind of Science*; Wolfram Media: Champaign, IL, USA, 2002; ISBN 1-57955-008-8.
13. Heinonen, T.; Pukkala, T. The use of cellular automaton approach in forest planning. *Can. J. For. Res.* **2007**, *37*, 2188–2200. [CrossRef]
14. Mathey, A.H.; Krcmar, E.; Tait, D.; Vertinsky, I.; Innes, J. Forest planning using co-evolutionary cellular automata. *For. Ecol. Manag.* **2007**, *239*, 45–56. [CrossRef]
15. Zeng, H.; Pukkala, T.; Peltola, H.; Kellomäki, S. Optimization of irregular-grid cellular automata and application in risk management of wind damage in forest planning. *Can. J. For. Res.* **2010**, *40*, 1064–1075. [CrossRef]
16. Strange, N.; Meilby, H.; Bogetoft, P. Land use optimization using self-organizing algorithms. *Nat. Resour. Model.* **2001**, *14*, 541–574. [CrossRef]

17. Strange, N.; Meilby, H.; Thorsen, J.T. Optimizing land use in afforestation areas using evolutionary self-organization. *For. Sci.* **2002**, *48*, 543–555.
18. Möykkynen, T.; Capretti, P.; Pukkala, T. Modelling the potential spread of Fusarium circinatum, the causal agent of pitch canker in Europe. *Ann. For. Sci.* **2015**, *72*, 169–181. [CrossRef]
19. Möykkynen, T.; Fraser, S.; Woodward, S.; Pukkala, T. Modelling of the spread of Dothistroma septosporum in Europe. *For. Pathol.* **2017**, *47*, 1–14. [CrossRef]
20. Möykkynen, T.; Pukkala, T. Modelling the spread of a potential invasive pest, the Siberian moth (Dendrolimus sibiricus) in Europe. *For. Ecosyst.* **2014**, *1*, 10. [CrossRef]
21. Baatz, M.; Schäpe, A. Multiresolution Segmentation: An Optimization Approach for High Quality Multi-Scale Image Segmentation. In Proceedings of the Angewandte Geograpische Informationsverarbeitung XII. Beiträge zum AGIT Symposium, Salzburg, Austria, 22–23 June 2000; Strobl, J., Griesebner, G., Blaschke, T., Eds.; Herbert Wichmann: Karlsruhe, Germany, 2000; pp. 12–23.
22. He, Q.; Chen, E.; Cao, C.; Liu, Q.; Pang, Y. A Study of Forest Parameters Mapping Technique Using Airborne LIDAR Data. *Adv. Earth Sci.* **2009**, *24*, 748–755.
23. *Forest Management Inventory Report of Mengjiagang Forest Farm*; Forest Monitoring and Planning Institute of Heilongjiang Province: Harbin, China, 2016. (In Chinese)
24. Vauhkonen, J.; Tokola, T.; Packalen, P.; Maltamo, M. Identification of single-tree attributes using airborne laser scanning-based height, intensity and alpha shape metrics. *For. Sci.* **2009**, *55*, 37–47.