

Evaluating structural indices by reversing forest structural analysis

Arne Pommerening*

School of Agricultural and Forest Sciences, University of Wales, Bangor, Gwynedd LL57 2UW, Wales, UK

Abstract

It is widely acknowledged that spatial forest structure is a driving factor behind growth processes and that forest growth, in return, influences the structural composition of woodlands. Also any impact on forests is primarily a change of spatial forest structure. In the last few decades an impressive number of structural indices have been developed to quantify spatial forest structure and it has also been suggested that they can be used as surrogate measures for quantifying biodiversity [Pommerening, A., 2002. Approaches to quantifying forest structures. *Forestry* 75, 305–324]. Of particular interest in this regard is the development of a family of individual tree neighbourhood-based indices, which are measures of small-scale variations in tree positions, species and dimensions, developed by Gadow and Hui [Gadow, K.v., Hui, G., 2002. Characterising forest spatial structure and diversity. In: Bjoerk, L. (Ed.), *Proceedings of the IUFRO International workshop 'Sustainable forestry in temperate regions'*, Lund, Sweden, pp 20–30]. Especially when expressed as frequency distributions these indices offer valuable information on spatial woodland structure. An important element of appraising the merits of such indices is a detailed evaluation of their performance for a specified purpose. One possible evaluation path is based on the idea that a successful quantification of spatial forest structure should allow the analysis to be reversed and enable the synthesis of forest structure from the indices derived. This idea is investigated here with a simulation model that uses the concept of cellular automata combined with further development of an approach by Lewandowski and Gadow [Lewandowski, A., Gadow, K.v., 1997. Ein heuristischer Ansatz zur Reproduktion von Waldbeständen (A heuristic method for reproducing forest stands). *Allg. Forst- u. J.-Zeitung* 168, 170–174]. The rules according to which the spatial pattern of tree positions “grows” in the stand matrix are deduced directly from the distributions of the structural indices of the input data. Different combinations of indices are used to assess and simulate the structure of four sample stands. The results show that simulations using species specific distributions of indices and a limit to the number of neighbours used for index calculation to three or four neighbours are most successful at reconstructing the original stand structure. The specific sequence of simultaneous distributions of structural indices was not significantly superior to the use of marginal distributions. Contrary to the suggestion in Hui et al. [Hui, G.Y., Albert, M., Gadow, K.v., 1998. Das Umgebungsmaß als Parameter zur Nachbildung von Bestandesstrukturen (The diameter dominance as a parameter for simulating forest structure). *Forstwiss. Centralbl.* 117, 258–266] no significant trend could be detected with regards to the use of the diameter dominance (formula 5) versus the diameter differentiation (formula 4).

The artificial synthesis of forest structure is of particular importance to conservationists who wish to develop forest landscapes to create a particular habitat pattern in order to support or re-introduce rare animal species. The topic is also important for modellers who require individual tree coordinates as input data for simulation runs or visualisations, which are hard to obtain in forest practice.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Spatial stand structure; Structural indices; Evaluation; Neighbourhood-based indices; Simulating spatial stand; Marginal and simultaneous distributions; Cellular automata

1. Introduction

A proper understanding of spatial forest structure is one of the keys to the sustainable management of mixed uneven-aged forests. The growth of trees is a reaction to their spatial context and conversely the growth processes influence the spatial forest structure and all biotic and abiotic, including human, impacts

modify spatial forest structure. A good understanding of these dependencies and their quantification is crucial for the management of woodlands for economic as well as environmental purposes.

The simulation or synthesis of spatial forest structure is an important aspect of environmental planning. For example, if there is a strong correlation between a particular spatial forest structure and the abundance of a particular animal species it should be possible to synthesize this structure elsewhere or at least to quantify the difference between the existing structure and an ideal structure in order to create new habitats for this

* Tel.: +44 1248 382440; fax: +44 1248 354997.

E-mail address: arne.pommerening@bangor.ac.uk.

Table 1

Overview of the most important individual tree indices developed by the Göttingen group since 1992, their formulae and the corresponding publications

No.	Index (reference)	Formula	Where
1	Uniform angle index (Gadow et al., 1998; Hui and Gadow, 2002)	$W_i = \frac{1}{n} \sum_{j=1}^n v_j$	$v_j = \begin{cases} 1, & \alpha_j < \alpha_0 \\ 0, & \text{otherwise} \end{cases}$
2	Species mingling (Füldner, 1995; Aguirre et al., 2003)	$M_i = \frac{1}{n} \sum_{j=1}^n v_j$	$v_j = \begin{cases} 1, & \text{species}_j \neq \text{species}_i \\ 0, & \text{otherwise} \end{cases}$
3	DBH differentiation (1) (Füldner, 1995; Pommerening, 1997, 2002)	$T_{ij} = 1 - \sum_{j=1}^n \frac{\min(\text{DBH}_i, \text{DBH}_j)}{\max(\text{DBH}_i, \text{DBH}_j)}$	j is the 1st neighbour tree
4	DBH differentiation (2) (Füldner, 1995; Gadow, 1999)	$T_i = 1 - \frac{1}{n} \sum_{j=1}^n \frac{\min(\text{DBH}_i, \text{DBH}_j)}{\max(\text{DBH}_i, \text{DBH}_j)}$	$j = 1 \dots n$ neighbour trees
5	DBH dominance (1) (Hui et al., 1998)	$U_i = \frac{1}{n} \sum_{j=1}^n v_j$	$v_j = \begin{cases} 1, & \text{DBH}_j \geq \text{DBH}_i \\ 0, & \text{otherwise} \end{cases}$
6	DBH dominance (2) (Gadow and Hui, 2002; Aguirre et al., 2003)	$U_i = \frac{1}{n} \sum_{j=1}^n v_j$	$v_j = \begin{cases} 1, & \text{DBH}_i \geq \text{DBH}_j \\ 0, & \text{otherwise} \end{cases}$

n is the number of neighbour trees. All index values are distributed between 0 and 1.

species. There is also an increasing demand for spatial tree data sets for certain growth models, sampling simulators and visualisation software (Pommerening, 2000). However, such data are very rare and there is often a need to simulate the necessary tree positions.

Over the past few decades many indices quantifying spatial forest structure have been developed (Pommerening, 2002) though very little work has been undertaken to formally evaluate these new developments (Neumann and Starlinger, 2001). One possible method of evaluation is to examine the correlation between the indices and more direct measures of biodiversity (e.g. the abundance of a particular rare species). An example of such a study using the indices discussed here is Spanuth (1998). However, another approach is to assess their ability to synthesize spatial forest structure. This would provide a closed circle starting with the analysis of spatial forest structure as a system and finishing with the reconstruction of this system. Such an evaluation can be understood as an application of systems analysis in that through the process of data reconstruction it is possible to learn what the advantages and the shortcomings of structural indices are. This is the approach to evaluating structural indices that this paper is concerned with.

2. Methods and data

2.1. Quantifying spatial forest structure

According to Gadow (1999) and Pommerening (2002) α -diversity can be subdivided into the diversity of tree positions (e.g. formula 1 in Table 1), tree species diversity (e.g. formula 2 in Table 1) and the diversity of tree dimensions (e.g. formulae 3–6 in Table 1). The diversity of tree positions reflects, at a small scale, whether the pattern of tree locations is regular, clumped (clustered), random or some combination of these. Tree species diversity is concerned with the spatial arrangement of species while diversity of tree dimensions involves the spatial arrangement of, for example, diameters or heights. The fundamental idea of this study is that a set of indices, which covers all three aspects of α -diversity should be sufficient to quantitatively describe and simulate spatial forest structure. Most structural indices attempt to examine and explain only the horizontal spatial structure and it has to be assumed that vertical aspects are allometrically related to horizontal aspects.

Since 1992 a research group at the Institute of Forest Management of the University of Göttingen (Germany) has developed a family of individual tree indices which are

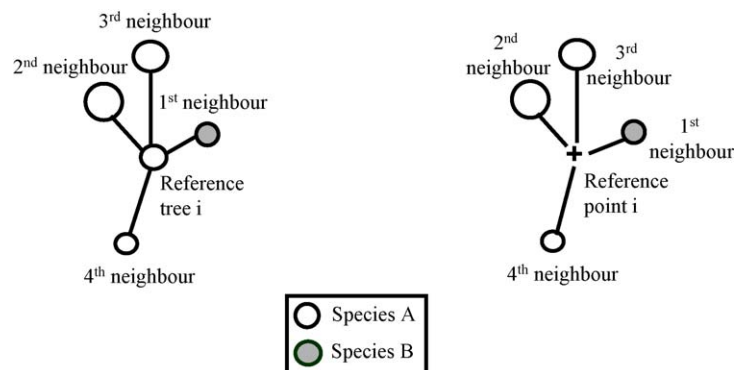


Fig. 1. Examples of a tree-based (left) and point-based (right) structural group involving in this case four neighbour trees. This structural group of reference tree/point and neighbouring trees is the computational unit for the indices listed in Table 1.

neighbourhood-based and which can account for small-scale aspects of the spatial distribution of tree attributes (Gadow and Hui, 2002; Fig. 1; Table 1). Their approach is similar to that of describing the structure of chemical molecules in that they focus on the quantification of the immediate neighbourhood of each tree or discrete points in a forest stand. An individual index value is then assigned to each tree, or to selected points, in the stand. Their algorithmic structure is very similar to that of distance dependent competition indices, which makes this family of indices particularly well suited to the simulation of spatial forest structure. Another advantage is that it is also comparatively easy to collect data for these indices during standard forest inventories (Füldner, 1995; Pommerening, 1997). The neighbourhood based indices are also very flexible in that the number of neighbour trees, n , to be considered is not fixed, except in the case of the uniform angle index. For this index $n = 4$ and a standard angle $\alpha_0 = 72^\circ$ ($=80$ gon) appear to be ideal (Hui and Gadow, 2002). The indices can be tree- or point-based (Gadow and Hui, 2002; Fig. 1) and, though point-based indices have certain advantages when used with inventories, only tree-based indices have been studied in this investigation because their concept is closer to that of cellular automata.

Tree-based indices result in an index value for each tree of the stand. An important development by the Göttingen research group was to make the distributions of index values the fundamental unit of information rather than the arithmetic mean values as in most of the previous approaches to constructing structural indices. Accordingly the main focus of this study is on the distributions of indices. With uniform angle index (formula 1), species mingling (formula 2) and DBH dominance (formulae 5 and 6) there is only a limited number of values the index can take. For example with $n = 4$ neighbours there are $n + 1 = 5$ possible values: 0.00, 0.25, 0.50, 0.75 and 1.00. Using these scores all trees of the stand can be classified and stand structure summarised by the distribution of values. With the indices mentioned above the number of classes is a function of the number of neighbours. This is not the case with the diameter differentiation (formulae 3 and 4) where a different approach to constructing classes has been chosen (see e.g. Pommerening, 2002).

2.2. A model frame work for evaluating structural indices

A simplified overview of the evaluation process is illustrated in Fig. 2. The object of interest is the forest, which is first measured, analysed and then re-synthesised using the indices described above (Table 1). Most investigations finish after the analysis stage but for evaluating the merit of structural indices a synthesis can be helpful. This would answer the question of how much can the indices contribute to a synthesis of forest structure. The synthesis is driven by a novel point process model, the probability distribution of which depends on the structural indices as kinds of super-parameters.

There are various approaches described in the literature for simulating spatial forest structure, which use classical models of the theory of point processes with a very small number of

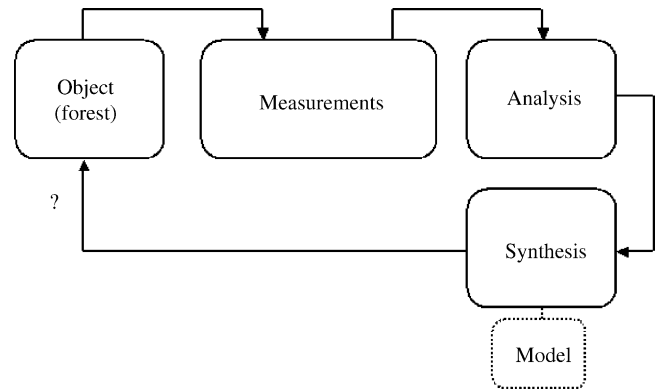


Fig. 2. Schematic overview of the evaluation process used in this study.

parameters; examples are Cox (Stoyan and Penttinen, 2000) or Gibbs processes (Tomppo, 1986). In their pure form most of these models are difficult to apply to forest structures with more than three species and their application is also limited when used with sample data. Therefore, more complicated point process models are necessary. Pretzsch (1997) developed an approach which is based on an empirical function, estimating the distance to the nearest neighbour, and a set of probability functions (a combination of an inhomogeneous Poisson process and a hard-core process) and Biber (1999) developed a method of simulating spatial forest structure in order to avoid edge effects in growth simulations. However, the latter two approaches are only applicable to forests that are located in the area where the data sets were taken for parameterization. Elaborating on Pretzsch's method, Pommerening (2000) generalised the empirical functions by deducing spatial information from forest inventories.

For an evaluation of the potential of the structural indices described in Table 1 for simulating forest structure it is necessary for the model to process the information generated by these indices only. The only approach close to the objectives of this study is that developed by Lewandowski and Gadow (1997) and parts of this are incorporated into the new method described below.

For this study a new class of point process models was developed using the concept of cellular automata. The method can be subdivided into three different phases (Fig. 3), with phase 1 modelling only the point pattern of tree positions, phase 2 modelling qualitative (tree species attributes) and phase 3 modelling quantitative marks (tree diameters). In phase 1 the spatial pattern of tree positions is simulated with the start configuration consisting of one tree being randomly placed on the stand grid. All the other $n - 1$ points are positioned subsequently using a set of rules of the type that drive cellular automata. More details about these techniques are provided in Section 2.3. The end product of phase 1 is a point pattern with no attributes assigned to the points other than coordinates, which are permanently fixed. The next two phases are an adaptation of the modelling approach published by Lewandowski and Gadow (1997). At the beginning of phase 2 the attribute 'tree species' is randomly assigned to all the points. This information can be obtained from a marginal species

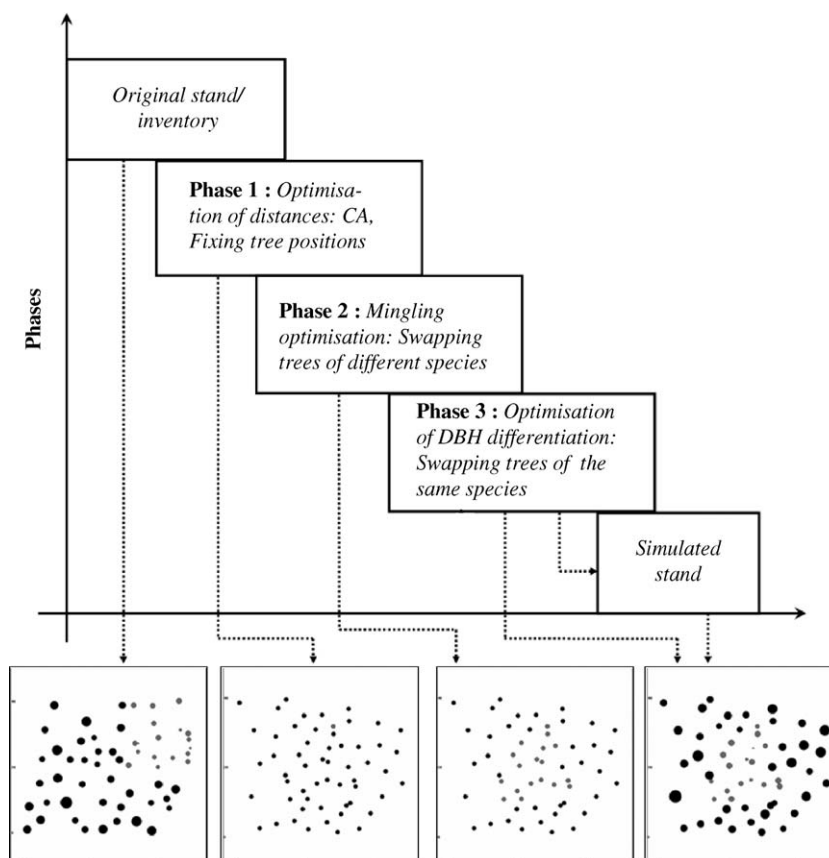


Fig. 3. The model sequence (CA: cellular automata, DBH: diameter at breast height).

distribution (no link with other distributions) or from a simultaneous distribution of species and uniform angle index. This random assignment is repeated 200 times and the configuration, which is closest to the empirical mingling distributions is carried forward to the optimisation part of this phase. During the optimisation of mingling the species attributes of all points with different attributes are swapped if this process of exchange results in a reduction of the deviation from the observed mingling distributions. Swapping trees of the same species would not change the mingling value. In each iteration all possible exchanges are investigated and only the one which results in the greatest reduction of the deviation from the observed mingling distributions is carried forward. When no further improvement is possible then phase 2 is finished. At the beginning of phase 3 diameters are assigned to all the points at random using diameter distributions with 1 cm-classes. The diameters can be assigned from marginal DBH distributions (non-species specific) or from simultaneous distributions of diameters and species. Phase 3 continues in a similar way to phase 2 but this time attributes (species and DBH) of tree locations are swapped only if they have the same species attribute. This restriction is necessary in order to retain the mingling configuration optimized in phase 2. Throughout this process all of the input tree attributes, coordinates, species and DBH are assigned to exactly the same number of trees as in the original stand or sample.

The last two phases of this modelling approach are similar to the idea of the simulated annealing reconstruction as described by Torquato (2002).

The model is not driven by pre-defined statistical functions, which would only be valid in a restricted geographic area. All the rules used in this approach are deduced from the input data which can come from full enumerations as well as inventories. It is, therefore, possible to collect the information required for calculating the structural indices as part of a standard forest inventory. Gadow et al. (2003) developed a method of estimating intertree distance distributions from the stand density and the mean of the uniform index value. This approach can be used to produce the necessary input data for this model if it is not feasible to sample tree distances in the field. However, the quality of the output does depend on the quality of the input and, the more accurate and unbiased the inventory is, the better the simulation result. In this study fully enumerated stand data were used in order to assess simulation errors. The structure and mechanism of the model is deliberately kept as simple as possible in order to retain a transparent process.

The data analysis methods, the model and most of the evaluation algorithms were implemented in an integrated software package called “CranCod” using structured programming techniques in the Borland Delphi environment which is based on the Pascal language. “CranCod” is currently being

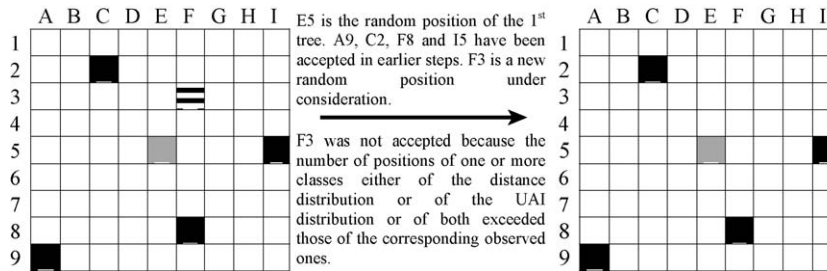


Fig. 4. An example of the working principle of the cellular automata used in this modelling approach.

revised and reorganised in an object-oriented way by using the language Java and the programming environment NetBeans.

2.3. Extended cellular automata

A cellular automaton (CA) is a system composed of cells, which are connected in regular structures. Each cell communicates its present state to its nearest neighbours and computes future states from its own, current, state and that of its neighbours. The concept of cellular automata was initially developed in the early 1950s by von Neumann and Ulam, and is an artificial life approach to simulation modelling. Since then the idea has been further developed by Conway in the 1970s, Wolfram in the 1980s (Wolfram, 2002) and others. CA have been successfully used for modelling biological systems for almost two decades (Ermentrout and Edelstein-Keshet, 1993; Ganguly et al., 2003) and can help to explain them especially where differential equations become too complex or do not

In a similar way the indices listed in Table 1 are based on small-scale neighbourhood relationships and therefore an application of cellular automata seems to be very appropriate as an approach to simulate tree positions.

The rules of the application of cellular automata of this model approach are derived from the intertree distance distribution to the first neighbour tree and the distribution of the uniform angle index. These two distributions were selected because Gadov et al. (2003) discovered a strong correlation between the uniform angle index (UAI, formula 1) and the distance to the first neighbour tree. The intertree distance distribution to the first neighbour tree is computed in a similar way to the diameter distribution by defining 50 cm distance classes¹. The cells of the simulated forest stand are examined and updated in random order as to whether they can accommodate a tree according to the distance and the UAI distributions. This simulation process is illustrated in Fig. 4.

The rule can be expressed in the following way (obs: observed, sim: simulated):

IF $number_{sim}[class[distance\ I]]+1 \leq number_{obs}[class[distance\ I]]$ **AND**
 $number_{sim}[class[UAI]]+1 \leq number_{obs}[class[UAI]]$ **THEN**
accept square mid-point as tree position
OTHERWISE
reject square mid-point as tree position

return a definite solution. There have been few applications of CA in forestry and most of these have dealt with the subjects of forest fires (e.g. Li and Magill, 2001; Karafyllidis and Thanailakis, 1997), land use and remote sensing (e.g. Soares-Filho et al., 2002).

Despite their simple construction cellular automata are able to simulate very complex behaviour and their application is very useful when studying systems consisting of many similar or identical units whose behaviour is influenced by local interactions. Characteristic of cellular automata is their ability for self-organisation. The computation laws for the new state are given by localized neighbourhood rules of occupancy (Zhang et al., 2004) and the behaviour of a CA can be specified by giving the structure of the interconnection grid, the neighbourhood of each cell, the boundary conditions, the rules for each cell in the grid and the initial conditions (Wolfram, 2002).

The distance distribution ensures that the simulated intertree distances do not fall below observed minimum distances and the UAI distribution ensures the correct geometric arrangement of tree positions. Unlike other approaches that simulate tree positions (e.g. Pretzsch, 1997) this method does not attempt to accommodate an array of trees with given attributes on the area of a forest stand. The simulation process starts with an array of available coordinates and subsequently examines whether they can host a tree or not.

The particular type of cellular automata used can be referred to as extended cellular automata. They deviate from standard CA in that the state of the two dimensional grid cells is updated

¹ The distance D1 between each tree of the forest stand and its first neighbour tree is allocated to a class of the distance distribution following the principle: $0\text{ m} \leq D1 < 0.5\text{ m}$, $0.5\text{ m} \leq D1 < 1\text{ m}$, $1\text{ m} \leq D1 < 1.5\text{ m}$, etc.

sequentially or asynchronously (Adachi et al., 2004). This basically means that every change in the state of a cell has an immediate effect on the state of all the other cells in the same time step. The simulation of the spatial tree pattern of a stand can therefore be completed in a single time step. The cells are examined and updated in random order using uniform random numbers either with or without replacement. Also the immediately adjoining cells do not define the neighbourhood as in earlier cellular automata applications (Wolfram, 2002). Rather the cells in which the tree neighbours are actually positioned (see Fig. 4) are used, and as these can vary the neighbourhood can be described as dynamic. The model and the algorithm for calculating the structural indices uses an edge-correction, which compensates for biased estimations caused by off-plot trees. In two-dimensional cellular automata the use of a so-called torus or toroidal wrapping (Diggle, 2003) is very common, in which the edges of the finite grid are pasted together resulting in a three-dimensional solid (Hegselmann and Flache, 1998; Diggle, 2003; Ripley, 1981). In forestry related modelling this method is also frequently referred to as “translation” (Monserud and Ek, 1974).

In this study the width of the interconnection grid in which the cellular automata operate is 5 cm.

2.4. Structural indices used in the different phases

The distributions of structural indices can be depicted either as marginal (univariate) or simultaneous (bivariate) and the model can work with both. Figs. 5 and 6 illustrate these different types of distributions.

The left diagram in Fig. 5 shows a marginal distribution of mingling (formula 2) regardless of species or any other index

while the right hand diagram offers us species-specific mingling distributions. The first two diagrams in Fig. 6 show marginal distributions of mingling (formula 2) and dominance (formula 5). The third diagram gives an impression of the simultaneous or bivariate distribution of mingling and dominance incorporating the information of the first two diagrams of Fig. 6.

Additionally, there is the option to work regardless of species or species specifically which can lead to simultaneous distributions of indices and species (i.e. a combination of the right hand diagrams in Figs. 5 and 6).

When developing rules for the structural synthesis the procedure varies depending on whether marginal or simultaneous distributions of the structural indices are used. For marginal distributions the procedure is as follows:

- Phase 1: Distance and UAI distributions (formula 1) sequentially
- Phase 2: Mingling distribution (formula 2)
- Phase 3: DBH differentiation (formula 3) or dominance distribution (formula 5)

In the case of simultaneous distributions of indices the model is organised in such a way that one index is always retained from one phase to the next to provide continuity throughout the simulation:

- Phase 1: Distance and UAI distributions simultaneously
- Phase 2: UAI and mingling distribution
- Phase 3: Mingling and DBH differentiation or dominance distribution

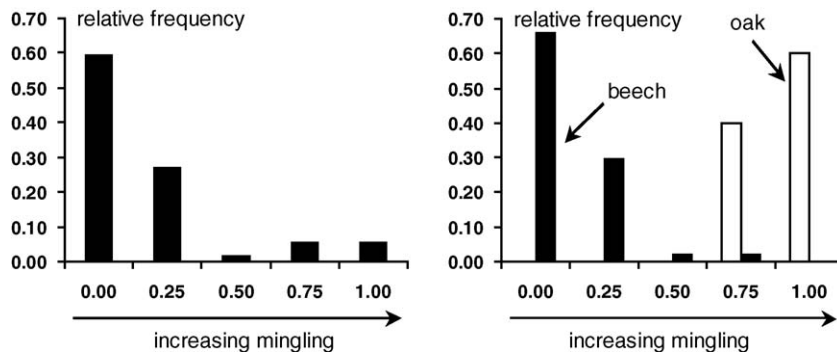


Fig. 5. The marginal mingling distribution (formula 2, left) and the simultaneous or bivariate distribution of mingling and species (right) for the German sample stand.

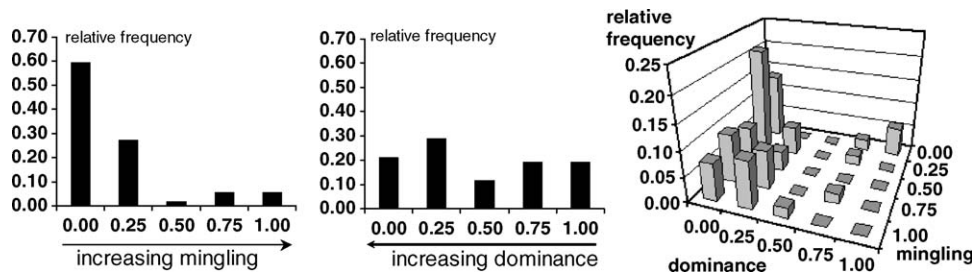


Fig. 6. The separate marginal distributions of mingling (formula 2) and dominance (formula 5, left) and the simultaneous or bivariate distribution of mingling and dominance (right) for the German sample stand.

By doing this it is hoped to retain the correlation between the structural indices throughout the simulation process.

The deviation between the observed and the simulated distributions of structural indices is measured with formula (7) using the absolute number of points/trees. The number of points/trees in each class of a distribution should be smaller or equal to the number of the corresponding observed one.

$$\text{Dev} = \sum_{i=1}^n |s_i - s'_i| \quad (7)$$

where n : number of classes, s_i : number of trees in class i of the observed distribution, s'_i : number of trees in class i of the simulated distribution.

The model is designed in such a way that each simulation run yields the exact basal area, mean squared diameter, stems per hectare and DBH distribution of the original forest.

2.5. Experimental plan and evaluation

The model is a tool for a multidimensional evaluation of structural indices and could use any number of factors to do this. However, because of the exponential increase of simulation runs required with increasing numbers of factors the experiment was restricted to five main factor groups:

1. The number of neighbouring trees: 3, 4, 5 and a mixed arrangement (four neighbours for the UAI distribution, three neighbours for the mingling distribution and three neighbours for the distributions of diameter differentiation/dominance).
2. DBH differentiation (formula 3) versus DBH dominance (formula 5) in phase 3.
3. Ordering of cells with/without replacement.
4. Marginal versus simultaneous distributions of structural indices.
5. Non-/species specific distributions of structural indices (phases 2 and 3 only).

This results in 64 possible different combinations of factors, each of which was simulated with five replications, giving a total of 320 model runs.

Three different groups of indices/functions have been used to evaluate the quality of the simulations. The first two groups comprise the indices which have been part of the simulation rules (formulae 1–3, 5, distance distribution) plus the additional index T (formula 4) and the unused index of the pair T1 (formula 3)/U (formula 5). The difference between group 1 and group 2 is that indices in the latter are species specific and in the former they are not. For group 2 the structural indices of the less abundant of the two main species was always selected. In the third group correlation functions were used for evaluation, i.e. pair-, mark correlation functions and mark connection functions (Stoyan and Penttinen, 2000; Pommerening, 2002). The pair correlation function only investigates the point pattern regardless of any other tree attributes while the mark correlation

function in this application also takes the tree diameters into account. The mark connection function calculates the probabilities that the tree species of a forest are associated with each other at particular intertree distances. In this study the intertree distances up to 10 m with a step width of 0.5 m were investigated. These continuous functions were used rather than indices like Clark and Evans (1954) or Pielou (1977) because they are more precise and deliver more information about the long-range structure of the simulated forests (Stoyan and Penttinen, 2000).

The main criterion to quantify the disparity between observed and simulated distributions in this application is the absolute discrepancy (AD; Gregorius, 1974; Pommerening, 1997; Aguirre et al., 2003).

$$\text{AD} = \frac{1}{2} \sum_{i=1}^n |s_i - s'_i| \quad \text{AD} \in [0, 1] \quad (8)$$

where n : number of classes, s_i : relative frequency of trees in class i of the observed distribution, s'_i : relative frequency of trees in class i of the simulated distribution.

It is defined as the relative proportion of individuals, which must be exchanged between the classes if the simulated distribution is to be transformed into the observed. Correspondingly, $1 - \text{AD}$ is the proportion common to both distributions. A value of $\text{AD} = 1$ means that both distributions have no common class, whereas $\text{AD} = 0$ signifies that the distributions are absolutely identical. AD is applied to all structural indices and to the mark connection function. The classes i are those of the indices as described in Section 2.1. As the mark connection function calculates probabilities for each species combination of a forest stand the probabilities at each discrete distance add up to 1 and can be interpreted as a class. The difference between observed and simulated pair and mark correlation functions are calculated as absolute differences similar to formula (7) because formula (8) is not applicable.

Because of the comparatively large number of individual simulations (320) and the evaluation criteria involved the process of identifying the best combination of factors had to be automated and simplified. For each of the three groups of evaluation criteria as described above arithmetic mean values were produced. In the case of the correlation functions this first required the calculation of an arithmetic mean over all intertree distances (from 0.5 to 10 m) for each individual simulation. Finally a total arithmetic mean was calculated by averaging the results of the three evaluation groups. The combination of factors with the lowest final mean value was identified as the best option. Additionally, the five factors as listed above were investigated individually.

2.6. Sample forest stand data

Four sample forest stands were selected to evaluate the performance of the structural indices. In order to underpin the universal applicability of the model the sample data sets were

Table 2

Basic stand data of the four sample forest stands used in this study (B/A: basal area, Dg: mean squared diameter, Clark & Evans: aggregation index of Clark and Evans (1954), Pielou: Pielou's (1977) coefficient of segregation)

Species	B/A (m ²)	Trees (per ha)	Dg (cm)
Germany (50 m × 50 m, 52 tree; oak 151, beech 65 years)			
Beech	21.4	752	19.0
Oak	21.3	80	58.3
Total	42.7	832	25.6
Clark & Evans	1.123	Pielou	−0.078
Greece (40 m × 30 m, 139 trees, fir 102, beech 166 years)			
Beech	34.2	758	24.0
Fir	8.3	400	16.2
Total	42.5	1158	21.6
Clark & Evans	0.798	Pielou	0.045
	B/A (m ²)	Trees (per ha)	D (cm)
Wales 1 (50 m × 40 m, 56 trees, both species 51 years)			
Sitka spruce	22.7	190	39.0
Lodgepole pine	3.9	90	23.5
–	–	–	–
Total	26.6	280	34.8
Clark & Evans	1.370	Pielou	0.836
Wales 2 (22 m × 12 m, 67 tree, all species 33 years)			
Sitka spruce	18.4	833	16.8
Birch	6.0	1288	7.7
Goat willow	1.5	417	6.7
Total	25.9	2538	11.4
Clark & Evans	0.989	Pielou	0.071

taken from very different European regions involving different species, environmental conditions, intervention history, etc.

Basic stand data including plot size, age, yield data and total number of trees are given in Table 2.

The stands are a mixed oak-beech stand (*Quercus robur* L. and *Fagus sylvatica* L.) from West Germany (Boeselager estate in the Sauerland region; Bölsing, 1996), a mixed beech-fir stand (*Fagus sylvatica* L./*Fagus moesiaca* Czeczott and *Abies borisii-regis* Matf.) from North Greece (Northern Pindos mountains; Beinsen, 1996) and two stands from North Wales (mixed Sitka spruce [*Picea sitchensis* (BONG.) CARR.]–lodgepole pine [*Pinus contorta* DOUGL. ex LOUD.] from Clocaenog forest and a mixed Sitka spruce–birch [*Betula pendula* ROTH]–goat willow [*Salix caprea* L.] site from Coed y Brenin).

The data cover a wide range of age and tree dimensions. According to the aggregation index of Clark and Evans (1954) the pattern of tree positions ranges from slightly clumped in

“Greece” to quite regular in “Wales 1”. “Wales 2” exhibits an almost random arrangement of tree positions. The pattern of tree species aggregation/segregation according to Pielou's (1977) coefficient of segregation ranges from close to randomness of species distribution in “Germany”, “Greece” and “Wales 2” to a segregation of species in “Wales 1”.

3. Results

Tables 3 and 4 show the combination of factors leading to the best and worst results from the four sample stands. The five factors are in the same order as listed in Section 2.5. The value in the column “deviation” is the mean of the absolute discrepancies (formula 8) of all evaluated indices/functions with the exception of the pair and the mark correlation functions where as stated earlier the absolute deviations have been used.

Although there is no overall consistency in the combinations that lead to the best or worst results it is possible to identify some trends on which hypotheses can be based. The worst results are generally linked to combinations where more than three or four neighbours are used for the calculation of structural indices. There does not seem to be any conclusive evidence as to whether the DBH differentiation (formula 1) or the DBH dominance (formula 5) is associated with better or worse results. The majority of best results are associated with an ordering of cells with replacement while the majority of worst results are associated with an ordering of cells without replacement. There is some uncertainty regarding the use of marginal or simultaneous distributions although all worst results are associated with simultaneous distributions. A clearer idea is conveyed with regard to non-species/species specific simulations. All the best options have been simulated with species specific distributions in phases 2 and 3 and all the worst ones with non-species specific ones apart from “Greece”, which seems to be an anomaly as will be demonstrated later. The differences between lowest and largest deviations are comparatively small which suggests that the indices show a similar simulation behaviour.

To further test the best choice of factors within groups 2–5, these were investigated using a paired *t*-test. For this purpose all the 64 results of the individual simulations were grouped in eight blocks with specific combinations of neighbours and DBH differentiation/dominance. However, for the investigation of DBH differentiation/dominance this was impractical and four blocks were identified. For each test all factors were kept constant except for the one under investigation. The

Table 3

The combination of factors leading to the best results (lowest deviation) with regard to the four sample stands

Sample stand	Neighbour combination	Differentiation or dominance	With or without replacement	Marginal or simultaneous	Non-species or species specific	Deviation
Germany	4 + 3 + 3	Dominance	Without	Simultaneous	Specific	0.1282
Greece	4 + 3 + 3	Differentiation	With	Simultaneous	Specific	0.0816
Wales 1	4 + 3 + 3	Differentiation	With	Marginal	Specific	0.1006
Wales 2	3 + 3 + 3	Differentiation	With	Marginal	Specific	0.0906

Table 4
The combination of factors leading to the worst results (highest deviation) with regard to the four sample stands

Sample stand	Neighbour combination	Differentiation or dominance	With or without replacement	Marginal or simultaneous	Non-species or species specific	Deviation
Germany	5 + 5 + 5	Differentiation	Without	Simultaneous	Non-specific	0.3026
Greece	3 + 3 + 3	Dominance	Without	Simultaneous	Specific	0.3059
Wales 1	5 + 5 + 5	Differentiation	Without	Simultaneous	Non-specific	0.2139
Wales 2	4 + 4 + 4	Dominance	With	Simultaneous	Non-specific	0.1908

hypothesized difference between the deviation means of alternative factors was set to zero. The results of the test show that there are only a few cases of significant difference between the use of factor options. Obviously there can be no definite statement as to whether the use of formula 3 is superior to the use of formula 5 or vice versa. The same is true for the question of whether an ordering of cells with replacement is advantageous in comparison with an ordering of cells without replacement. With regard to the superiority of marginal or simultaneous distributions of structural indices there is only one significant case in “Wales 1” where the use of simultaneous distributions is more advantageous. This is the stand with a segregation of the two species, which are correlated with very different diameter ranges. With the majority of sample stands the use of species-specific distributions of structural indices in phases 2 and 3 is significantly superior to the use of non-species specific ones. As detected in Tables 4 and 5 there is a somewhat different simulation behaviour when using the Greek sample data, with results generally not following the pattern of the other three.

In order to investigate the performance of different combinations of neighbours a single factor ANOVA test was carried out and the results are shown in Table 6. It is apparent that in only two cases are at least two of the four neighbour combinations significantly different at the 0.05 level.

A consequential pairwise analysis of the significant cases in Table 6 using the method described in Bortz (1999, p. 252f) reveals that with “Wales 1” simulations using three neighbours (3 + 3 + 3) are superior to those using four neighbours (4 + 4 + 4). The analysis also shows that the mixed scenario (4 + 3 + 3) is significantly superior to the use of five (5 + 5 + 5) and four neighbours (4 + 4 + 4). With the “Greece” data set the relation between different sets of numbers of neighbours is to the contrary. Using four neighbours (4 + 4 + 4) is superior to using three neighbours (3 + 3 + 3), the mixed scenario (4 + 3 + 3) and five neighbours (5 + 5 + 5). Other comparisons are not significant.

Table 5
Results of the paired *t*-test for the factors 2–5

Sample stand	Differentiation vs. dominance $t_{(3;0.05)} = 3.18$	With vs without replacement $t_{(7;0.05)} = 2.36$	Marginal vs. simultaneous $t_{(7;0.05)} = 2.36$	Non-species vs. species specific $t_{(7;0.05)} = 2.36$
Germany	1.6115	−1.8340	−1.3870	8.7286 ^a
Greece	1.8748	−0.4574	−1.0936	0.8779
Wales 1	−1.7160	1.4215	−4.8292 ^a	3.9985 ^a
Wales 2	−1.1990	0.4242	−1.7377	12.6984 ^a

^a Deviation means are significantly different at the 0.05 level [2-tailed].

To illustrate the results a visualisation of the best and the worst individual simulation results for “Wales 1” (see Tables 3 and 4) in relation to the original stand structure is shown in Fig. 7. In both cases a visual impression suggests that the species segregation has been simulated reasonably well. As the model uses an edge correction all the lodgepole pines in the top right corner of the original stand are surrounded by Sitka spruces in the two simulations. However, in the worst simulation lodgepole pine trees are given much larger diameters than in the original stand. This is a typical effect of non-species specific distributions of structural indices in phases 2 and 3 because information regarding species and

Table 6
Results of the single factor ANOVA test for the factor 1

Source of variation	SS	d.f.	MS	<i>F</i>	<i>F</i> _{crit}
Germany					
Between groups	0.0134	3	0.0045	2.75	2.95
Within groups	0.0455	28	0.0016		
Total	0.0589	31			
Greece					
Between groups	0.0770	3	0.0257	13.68 ^a	2.95
Within groups	0.0525	28	0.0019		
Total	0.1296	31			
Wales 1					
Between groups	0.0067	3	0.0022	5.32 ^a	2.95
Within groups	0.0117	28	0.0004		
Total	0.0183	31			
Wales 2					
Between groups	0.0015	3	0.0005	0.63	2.95
Within groups	0.0220	28	0.0008		
Total	0.0235	31			

SS: sum of squares, d.f.: degrees of freedom, MS: mean square, *F*: empirical *F*-value, *F*_{crit}: critical *F*-value.

^a Deviation means are significantly different at the 0.05 level [two-tailed].

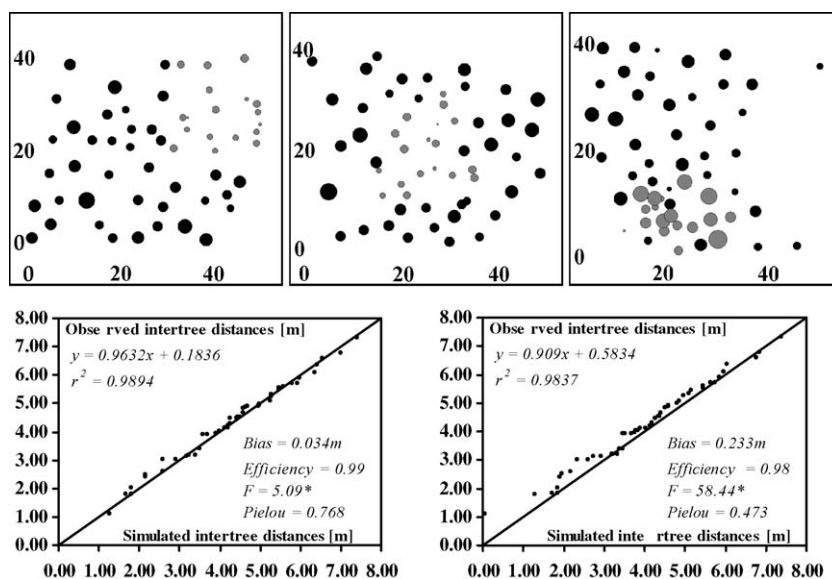


Fig. 7. Tree location maps and evaluation graphs of the original data set (top left), the first replication of the optimal simulation (top centre) and of the worst simulation (top right) for the sample stand “Wales 1” (Black: Sitka spruce, Grey: Lodgepole pine). The graphs showing the statistics bias, efficiency and simultaneous F -test were calculated according to Sterba et al. (2001) and refer to the best and the worst simulation.

dimension is separated. The statistical figures and the graphs suggest that the point patterns of all three data sets are not so different which is the result of the cellular automata approach. However, the bias of the worst simulation is much larger than that of the best one. While the Pielou (1977) value of the best simulation is reasonably close to that of the original stand (Table 1), the cluster effect in the worst simulation is exaggerated.

4. Discussion and conclusions

This multidimensional approach to evaluating structural indices has highlighted a number of valuable aspects as to how structural indices work and the degree to which they are able to quantify spatial forest structure. The simulation results support Wolfram’s (2002) argument that simple rules such as distributions of spatial indices can lead to quite complex point patterns like the spatial arrangement of two-dimensional tree positions.

Although there do not seem to be any overall trends in the five factors investigated there are a few aspects that deserve further study in the future. There appears to be an advantage in using species-specific distributions especially in stands with segregated species. The use of more than five neighbours and in some cases also four neighbours for calculating and simulating the three aspects of α -diversity are not optimal. This finding is helpful with regard to sampling and quantifying the indices of Table 1 as part of forest inventories based on circular sample plots where a requirement for a larger number of neighbours can lead to a significant bias due to edge effects (Pommerening and Gadow, 2000). Hui et al. (1998) state that the use of the diameter dominance is superior to the use of the diameter differentiation when simulating spatial forest structure but this could not be confirmed in this study. The fact that there is no trend concerning the use of marginal versus simultaneous

distributions could be explained by a lack of understanding of the correlations between structural indices. These correlations seem to be specific to different forest stands or at least to groups of different forest stands. Instead of utilising a rigid chain of simultaneous distributions as in this paper (see Section 2.4) it might be better to identify the specific correlations for each stand and then develop the chain of simultaneous distributions accordingly. The comparatively good performance of marginal distributions can also be explained by the fact that these have fewer classes that are not so specific.

The point-based versions of the indices (see right hand diagram in Fig. 1; Gadow and Hui, 2002) used in this study could, theoretically, also be used with the cellular automata approach although the tree-based concept is closer to the original idea of CA. The total population of point-based indices, however, is based on all possible points within the stand boundaries, which is infinite. The total population of tree-based indices consists only of those points within the stand boundaries, which are tree positions. The rule of the CA would therefore need to be adapted. With tree-based indices each newly accepted point represents a reference tree and potentially a neighbour of other reference trees. With point-based indices each newly accepted point represents a neighbour tree of points only. An application of cellular automata to simulate the pattern of tree positions could be achieved by defining the mid points of all the cells as the total population of points to be examined in the analysis and the synthesis.

The only other indices, which are similar to those used in this study are distance dependent competition indices (Biging and Dobbertin, 1992). Although their main concern is to quantify competition pressure for each tree of a forest stand it might be possible to test the merits of some competition indices, or to merge some aspects of their concepts, with those of the structural indices of Table 1. The similarity between structural and competition indices is particularly obvious with the

dominance index (formulae 5 and 6 in Table 1). It would also be interesting to explore whether the inclusion of a vertical dimension could improve the simulation of the two-dimensional patterns of tree positions.

As mentioned in Section 2.2 the use of cellular automata is only one possible approach to simulating spatial patterns and can be understood as a special case of modelling spatial correlation. In fact, this method is not so different from the idea of Gibbs processes.

As a positive by-product of this evaluation study it has proved possible to optimise a model for simulating spatial forest structure which is not based on pre-defined statistical functions and can be readily applied anywhere in the world without the need for local adaptation. Although this approach also uses quite a few empirical parameters, i.e. the structural indices, these are individually deduced from the input data for each simulation. Given the demand for spatially explicit individual tree data such a model could become an important tool for other investigations that need such data and the modelling process itself will help to develop a better understanding of spatial forest structure. Similar index approaches at landscape level could allow the simulation of larger entities.

Acknowledgements

The author wishes to thank the Welsh European Funding Office and the Forestry Commission Wales for the financial support for the Tyfiant Coed project of which this study is a part. Gratitude is also expressed to the Research Committee of the School of Agricultural and Forest Sciences of the University of Wales, Bangor for financial and moral support. This work has been very much stimulated by discussions with colleagues of the EFI Project Centre CONFOREST. I would also like to thank my colleagues Hubert Sterba, Steve Murphy, Jens Haufe, Owen Davies and two anonymous referees for helpful comments on earlier drafts of the text.

References

- Adachi, S., Peper, F., Lee, J., 2004. Computation by asynchronously updating cellular automata. *J. Stat. Phys.* 114, 261–289.
- Aguirre, O., Hui, G.Y., Gadow, K., Jiménez, J., 2003. An analysis of spatial forest structure using neighbourhood-based variables. *For. Ecol. Manage.* 183, 137–145.
- Beinsen, H., 1996. Inventur und Strukturbeschreibung von zwei Naturwaldzellen im Pindos-Gebirge, Griechenland [Inventory and structural analysis of two monitoring plots in the Pindos mountains, Greece]. Diploma thesis, Göttingen Forestry School, 103 pp.
- Biber, P., 1999. Ein Verfahren zum Ausgleich von Randeffecten bei der Berechnung von Konkurrenzindizes [A method for edge-bias compensation when calculating competition indices]. In: Kenk, G. (Ed.), 1999: Deutscher Verband Forstlicher Versuchsanstalten. Sektion Ertragskunde. Proceedings of the Jahrestagung 1999. Freiburg, pp.189–202.
- Biging, G.S., Dobbertin, M., 1992. A comparison of distance-dependent competition measures for height and basal area growth of individual conifer trees. *For. Sci.* 38, 695–720.
- Bölsing, S., 1996. Strukturbeschreibung von vier Eichenbeständen unterschiedlichen Alters [Structural analysis of four oak stands of different age]. Diploma thesis, University of Göttingen, 75 pp.
- Bortz, J., 1999. Statistik für Sozialwissenschaftler [Statistics for Social Scientists], fifth ed. Springer Verlag, Berlin, p. 836.
- Clark, Ph. J., Evans, F.C., 1954. Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology* 35, 445–453.
- Diggle, P.J., 2003. Statistical Analysis of Spatial Point Patterns, second ed. Arnold, London, p. 159.
- Ermentrout, G.B., Edelstein-Keshet, L., 1993. Cellular automata approaches to biological modelling. *J. Theor. Biol.* 160, 97–133.
- Füldner, K., 1995. Strukturbeschreibung von Buchen-Edellaubholz-Mischwäldern [Describing forest structures in mixed beech-ash-maple-sycamore stands]. Ph.D. dissertation, University of Göttingen, Cuvillier Verlag, Göttingen, 163 pp.
- Gadow, K.v., Hui, G.Y., Albert, M., 1998. Das Winkelmaß—ein Strukturparameter zur Beschreibung der Individualverteilung in Waldbeständen [The uniform angle index—a structural parameter for describing tree distribution in forest stands]. *Centralbl. ges. Forstwesen* 115, 1–10.
- Gadow, K.v., Hui, G.Y., Chen, B.W., Albert, M., 2003. Beziehungen zwischen Winkelmaß und Baumabständen [Relationship between the uniform angle index and nearest neighbour distances]. *Forstwiss. Centralbl.* 122, 127–137.
- Gadow, K.v., Hui, G., 2002. Characterising forest spatial structure and diversity. In: Bjoerk, L. (Ed.), Proceedings of the IUFRO International workshop ‘Sustainable forestry in temperate regions’, Lund, Sweden, pp. 20–30.
- Gadow, K.v., 1999. Waldstruktur und Diversität [Forest structure and diversity]. *Allg. Forst- u. J. -Zeitung* 170, 117–122.
- Ganguly, N., Sikdar, B.K., Deutsch, A., Canright, G., Chaudhuri, P.P., 2003. A survey on cellular automata. Technical Report Centre for High Performance Computing, Dresden University of Technology, 30 p. <http://www.cs.uni-bo.it/bison/publications/CAsurvey.pdf>.
- Gregorius, H.-R., 1974. Genetischer Abstand zwischen Populationen. I. Zur Konzeption der genetischen Abstandsmessung [Genetic distance among populations. I. Towards a concept of genetic distance measurement] *Silvae Genet.* 23, 22–27.
- Hegselmann, R., Flache, A., 1998. Understanding complex social dynamics: a plea for cellular automata based modelling. *J. Artif. Soc. Soc. Simulat.* 1, In: <http://jasss.soc.surrey.ac.uk/1/3/1.html>.
- Hui, G.Y., Albert, M., Gadow, K.v., 1998. Das Umgebungsmaß als Parameter zur Nachbildung von Bestandesstrukturen [The diameter dominance as a parameter for simulating forest structure]. *Forstwiss. Centralbl.* 117, 258–266.
- Hui, G.Y., Gadow, K.v., 2002. Das Winkelmaß. Herleitung des optimalen Standardwinkels. [The uniform angle index. Derivation of the optimal standard angle] *Allg. Forst- u. J. -Zeitung* 173, 173–177.
- Karafyllidis, I., Thanailakis, A., 1997. A model for predicting forest fire spreading using cellular automata. *Ecol. Model.* 99, 87–97.
- Lewandowski, A., Gadow, K.v., 1997. Ein heuristischer Ansatz zur Reproduktion von Waldbeständen [A heuristic method for reproducing forest stands]. *Allg. Forst- u. J. -Zeitung* 168, 170–174.
- Li, X., Magill, W., 2001. Modelling fire spread under environmental influence using a cellular automaton approach. *Complex. Int.* 8, 1–14.
- Monserud, R.A., Ek, A.R., 1974. Plot edge bias in forest stand growth simulation models. *Can. J. For. Res.* 4, 419–423.
- Neumann, M., Starlinger, F., 2001. The significance of different indices for stand structure and diversity in forests. *For. Ecol. Manage.* 145, 91–106.
- Pielou, E.C., 1977. *Mathematical Ecology*. John Wiley and Sons, p. 385.
- Pommerening, A., 1997. Eine Analyse neuer Ansätze zur Bestandesinventur in strukturreichen Wäldern [An analysis of new approaches towards stand inventory in structure-rich forests]. Ph.D. dissertation, Faculty of Forestry and Forest Ecology, University of Göttingen, Cuvillier Verlag Göttingen, 187 pp.
- Pommerening, A., 2000. Neue Methoden zur räumlichen Reproduktion von Waldbeständen und ihre Bedeutung für forstliche Inventuren und deren Fortschreibung [New methods of spatial simulation of forest structures and their implications for updating forest inventories]. *Allg. Forst- u. J. -Zeitung* 171, 164–169.
- Pommerening, A., 2002. Approaches to quantifying forest structures. *Forestry* 75, 305–324.

- Pommerening, A., Gadow, K.v., 2000. Zu den Möglichkeiten und Grenzen der Struktur erfassung mit Waldinventuren [On the possibilities and limitations of quantifying forest structures by forest inventories]. *Forst und Holz* 55, 622–631.
- Pretzsch, H., 1997. Analysis and modelling of spatial stand structures. Methodological considerations based on mixed beech-larch stands in Lower Saxony. *For. Ecol. Manage.* 97, 237–253.
- Ripley, B.D., 1981. *Spatial Statistics*. John Wiley & Sons, New York, p. 252.
- Soares-Filho, B.S., Cerqueira, G.C., Pennachin, C.L., 2002. Dinamica—a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecol. Model.* 154, 217–235.
- Spanuth, M., 1998. Untersuchungen zu den Hauptschlafbaumarten Eiche, Fichte und Buche des Waschbären (*Procyon lotor*) im südlichen Solling. [Investigation of the main tree species oak, spruce and beech preferred by racoons (*Procyon lotor*) as resting places in the southern Solling]. Diploma thesis, University of Göttingen, 64 pp.
- Sterba, H., Korol, N., Rössler, G., 2001. Ein Ansatz zur Evaluierung eines Einzelbaumwachstumssimulators für Fichtenreinbestände [An approach to evaluating an individual tree growth simulator for pure spruce stands]. *Forstwiss. Centralbl.* 120, 406–421.
- Stoyan, D., Penttinen, A., 2000. Recent applications of point process methods in forestry statistics. *Stat. Sci.* 15, 61–78.
- Tomppo, E., 1986. *Models and Methods for Analysing Spatial Patterns of Trees*, vol. 138. *Communicationes Instituti Forestalis Fennica*, Helsinki, p. 65.
- Torquato, S., 2002. Random heterogeneous materials. In: *Microstructure and Macroscopic Properties*, Springer Verlag, p. 701.
- Wolfram, S., 2002. *A New Kind of Science*, first ed. Wolfram Media Inc., p. 1197.
- Zhang, J., Wang, H., Li, P., 2004. Cellular automata modelling of pedestrian's crossing dynamics. *J. Zhejiang Univ. Sci.* 5, 835–840.