

Edge-correction needs in estimating indices of spatial forest structure

Arne Pommerening and Dietrich Stoyan

Abstract: Indices quantifying spatial forest structure are frequently used to monitor spatial aspects of tree attributes including biodiversity in research plots of limited size. The treatment of edge trees, which are close to the plot boundaries, can affect the estimation of such indices that include neighbour effects, since some of their neighbours are likely to fall outside the plot. This paper investigates whether and under what circumstances edge-correction methods are necessary and evaluates the performance of six different approaches: no edge correction, translation, reflection, buffer zone, and two new nearest-neighbour methods. The performance of edge-correction methods depends strongly on the algorithmic structure of the indices and the spatial pattern of tree positions involved. Some edge-correction methods introduce more error than ignoring edge bias altogether. For indices accounting for the diversity of tree positions and especially for those computing angles, translation or buffer zone methods reduce the estimation error regardless of the sample size. The use of the reflection method is associated with large bias values. One of the new nearest-neighbour edge-correction methods proves to be capable of reducing the bias considerably. The results confirm the need for sufficiently large monitoring plots to avoid bias from edge effects. Where this is impossible, neighbours beyond the plot boundary need to be included in the survey, thus providing unbiased estimates but at the cost of extra measurements. Sensitivity analysis is required for newly introduced indices prior to their first application.

Résumé : Des indices qui permettent de quantifier la structure spatiale de la forêt sont fréquemment utilisés pour faire le suivi de l'aspect spatial des attributs des arbres, incluant la biodiversité dans les placettes expérimentales à superficie restreinte. Le traitement des arbres situés en bordure de la placette peut affecter l'estimation de ces indices puisque certains de leurs voisins peuvent vraisemblablement se retrouver à l'extérieur de la placette. Cet article examine si des méthodes de correction des effets de bordure sont nécessaires et dans quelles circonstances. Les auteurs analysent également la performance de six approches différentes : aucune correction de l'effet de bordure, translation, réflexion, zone tampon et deux nouvelles méthodes du plus proche voisin. La performance des méthodes de correction est fortement dépendante de la structure de l'algorithme des indices et de la configuration spatiale de la position des arbres concernés. Certaines méthodes de correction introduisent une erreur plus importante que le fait de simplement ignorer les biais dus à l'effet de bordure. Pour le groupe d'indices qui tiennent compte de la diversité dans la position des arbres et spécialement pour ceux qui servent au calcul des angles, les méthodes de translation et de la zone tampon réduisent l'erreur d'estimation sans égard à la taille de l'échantillon. La méthode par réflexion génère des biais élevés. L'une des deux nouvelles méthodes du plus proche voisin a pu considérablement réduire le biais. Les résultats confirment la nécessité que les placettes aient une superficie suffisamment grande pour éviter les biais dus aux effets de bordure. Lorsque cela est possible, les voisins situés à l'extérieur des limites de la placette devraient être inclus dans l'inventaire pour obtenir des estimations non biaisées au prix, cependant, de mesures supplémentaires. Il est nécessaire de procéder à l'analyse de sensibilité des nouveaux indices avant de les utiliser pour la première fois.

[Traduit par la Rédaction]

Introduction

The quantitative characterization of spatial structure is an important part of the study of ecological processes in forests (Upton and Fingleton 1985, 1989; Pretzsch 1998; Barot and Gignoux 2003; Bauer et al. 2004). For this purpose many statistical summary characteristics have been developed that characterize forest structure and aspects of biodiversity using numerical values or functions. Examples of such functions are Ripley's K function, $K(r)$, or the nearest-neighbour dis-

tance distribution function, $G(r)$ (Stoyan and Stoyan 1994), while typical examples of numerical characteristics are structural indices such as the aggregation index (Clark and Evans 1954) and the species segregation index (Pielou 1977).

Usually it is assumed that the two-dimensional arrangement of tree positions in a forest can be described by a point process (Penttinen et al. 1992). In a point process each individual tree, i , can be understood as a point or event defined by its position in the monitoring plot or observation window W using two-dimensional Cartesian coordinates $\{x_i, y_i\}$. The

Received 3 November 2005. Accepted 2 March 2006. Published on the NRC Research Press Web site at <http://cjfr.nrc.ca> on 13 June 2006.

A. Pommerening,¹ School of Agricultural and Forest Sciences, University of Wales, Bangor, Gwynedd, LL57 2UW, Wales, UK.
D. Stoyan, Institut für Stochastik, TU Bergakademie Freiberg, D-09596 Freiberg, Germany.

¹Corresponding author (e-mail: arne.pommerening@bangor.ac.uk).

observation window is usually a rectangular or circular area in the forest that is assumed to yield representative information on the forest as a whole. Marks $\{m_{ij}, \dots, m_{in}\}$ are often attached to the points, which are sets of qualitative (e.g., species) and (or) quantitative (e.g., diameter at breast height, DBH) values. Frequently, these marks are constructed or computed, for example, competition indices describing the influence of neighbour trees on the reference tree. These indices are usually calculated by means of mathematical formulae using intertree distances and original tree marks such as DBH or species.

In statistical analysis it is assumed that the point process studied is spatially homogeneous (or, in terms of point process theory, stationary). Under the spatial homogeneity assumption it is meaningful to characterize the tree point pattern of the underlying process, in the research plot, using global characteristics. The simplest of these are perhaps mean number of stems per hectare (SPH), mean basal area, or mean volume per hectare. In ecological research other characteristics are used, such as the mean values of computed individual tree characteristics, for example, mean competition indices (Barot and Gignoux 2003; Bauer et al. 2004). Also, $K(r)$ and $G(r)$ are characteristics that have meaning only in the homogeneous case.

When calculating spatial characteristics, edge effects are likely to play an important role. Ignoring these effects is usually believed to result in biased statistical estimations. This is because some of the immediate neighbours of trees near the boundary of W are outside W and therefore not recorded; however, these trees are needed to calculate the correct values of statistical quantities such as competition or nearest neighbour indices. While edge-bias issues of correlation functions and competition indices have been frequently discussed in the statistical literature (e.g., Monserud and Ek 1974; Martin et al. 1977; Stoyan and Stoyan 1994; Stoyan et al. 1995; Radtke and Burkhart 1998; Gignoux et al. 1999), comparatively little attention has been paid to structural indices (Donnelly 1978). As the algorithms of nearest-neighbour structural indices are very similar to those of competition indices, it is likely that careless treatment of edge trees may also lead to a bias in the estimation of such indices especially when the window of observation is small and (or) tree numbers are low. Such biased estimation of structural indices can lead to incorrect conclusions and interpretations of the spatial biodiversity of a woodland. It should be noted that the situation is different with statistical tests, for example, concerning the hypothesis of complete spatial randomness, as in Gignoux et al. (1999). These authors compared summary characteristics of the Poisson process with those from its samples and found that tests without edge correction have a higher power for small sample sizes than those with correction.

The attention paid to edge effects has varied, and the early publications of Clark and Evans (1954) and Pielou (1977) simply ignored the problem. Later, Donnelly (1978) investigated the effect of edge bias on the aggregation index of Clark and Evans (1954) and developed an empirical edge correction specifically for this index. Since then various forms of edge correction have been developed. A very simple one, used by many researchers (e.g., Neumann and Starlinger 2001; Aguirre et al. 2003), is the use of a bound-

ary strip or "buffer" zone, of width d , inside the monitoring plot W . This method is also referred to as the border or guard method (Ripley 1981). In the statistical analysis only the trees in a reduced window of observation are used as reference trees, namely those that have a distance larger than d from the window boundary. The width of the buffer zone (e.g., $d = 5$ m) should be large enough that all relevant interaction between trees in the reduced window is accounted for within W . Determining the optimal width of the buffer zone is difficult; if it is too small residual edge effects will remain; if it is too large valuable data are discarded unnecessarily (Diggle 2003).

Other edge-correction methods can be considered as being only speculative (Pretzsch 2002; Diggle 2003), as they extrapolate the spatial structure from within W to an infinite plane and join parts of the point pattern that do not occur so close together in nature. Two examples are translation (also referred to as torus or periodic boundary conditions; Diggle 2003) and reflection (e.g., Radtke and Burkhart 1998; see Fig. 1).

It is clear that these two methods do not easily work with circular sample plots or with plots of irregular shape (Windhager 1997). Furthermore, translation and reflection are believed to result in unrealistic periodicities of the spatial point pattern and therefore neighbourhood situations that do not naturally occur (Pretzsch 2002; Diggle 2003).

This paper explores the need for edge correction of five spatial structural indices involving nearest neighbours. The performance of six edge-correction methods for these indices is investigated through simulations. Included in the study are two recently developed nearest-neighbour edge-correction concepts, NN1 and NN2 (Hanisch 1984).

Materials and methods

Indices used in this study

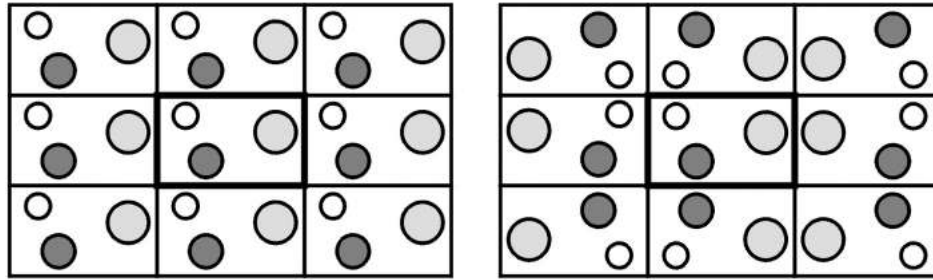
Tables 1 and 2 give an overview of the nearest-neighbour indices used in this study.

The integer variable n in Table 1 is the number of nearest neighbours included in the determination of the index. In this study, n is set to 4 for indices 1, 2, and 3, while the DBH differentiation, Clark and Evans, and Pielou indices generally use only one neighbour. The uniform angle index is derived from the angles subtended by adjacent neighbours at the reference trees. The six indices were selected so that similar aspects of spatial structure are accounted for by two indices (1 and 5, 2 and 6, 3 and 4; Table 1).

The two new edge-correction concepts, NN1 and NN2 (Hanisch 1984), can be applied to any structural index that is based on the nearest-neighbour principle, for example, those of Clark and Evans (1954) and Pielou (1977). The mingling index, an indicator of spatial species diversity (Füldner 1995; Fig. 2), will be used to explain and demonstrate NN1 and NN2. It is representative of a group of neighbourhood-based indices describing the spatial distribution of tree attributes.

The mingling index (M_i) gives, for each individual tree i , the proportion of its n nearest neighbours that do not belong to the same species as the reference tree i . Figure 2 illustrates the index for the case $n = 4$.

Fig. 1. Illustration of the translation (left) and reflection (right) edge-correction methods used in this study. The observation window, W , is in the centre with boundaries in bold.



The mingling index can only take one of a limited number of values. For example, when $n = 4$ neighbours there are $n + 1 = 5$ possible values of M_i , calculated as k/n , with $k = 0, 1, \dots, n$: 0.00, 0.25, 0.50, 0.75, and 1.00 (Fig. 2). Using these scores, all trees of the stand can be individually classified, and the overall stand structure can be characterized by the distribution of the values of M_i . The mean population mingling, M , can be expressed as

$$[1] \quad M = \sum_{k=1}^n \frac{k}{n} m_k = \frac{1}{N} \sum_{i=1}^N M_i$$

where m_k is the proportion of trees that have a mingling index of k/n and N is the number of trees in the population considered. Naturally, the full distribution of m_k provides more information than M .

The number n is fixed for all trees of an observation window, but can be chosen by the forestry statistician, determined by the objectives of individual studies on species diversity, the size of the window of observation and the density of points. For example, Fuldner (1995) recommended $n = 3$ neighbours, while more recent publications (Aguirre et al. 2003) have suggested that $n = 4$ neighbours is to be preferred. For more information on how to use and interpret the mingling index and the other indices listed in Table 1, see Aguirre et al. (2003) and Pommerening (2002).

NN1 and NN2 estimators

In addition to the existing edge-correction methods, which will be described in Table 3, the authors have considered two new concepts, the NN1 and NN2 estimators. Stoyan (2005) demonstrated that NN1 in particular, which has its origin in Hanisch (1984), is a very successful estimator of the nearest-neighbour distance distribution function, $G(r)$, and therefore it seemed appropriate to study its performance in the estimation of indices. The estimator NN1 is ratio-unbiased and includes more points than the unbiased buffer zone estimator, which, as noted earlier, has the additional disadvantage that buffer zone width is statistically difficult to determine. NN1 and the closely related NN2 are specifically designed for the statistical estimation of characteristics related to nearest neighbours and so are referred to as nearest-neighbour or NN-correction estimators. The principle is explained in Stoyan and Stoyan (1994, p. 297) for the case of $G(r)$. The idea is to use variable buffer zones for the trees close to the edge of the observation window. NN1 and NN2 incorporate a buffer but, in contrast to the traditional buffer method with a constant d , the decision to exclude trees close

to the window's edge is made individually, based on the spatial arrangement of the trees.

In the statistical estimation of nearest-neighbour indices, tree i is used as a reference tree only if its distance s_i to the boundary of the window W is further than or equal to its distance c_i from its n th nearest neighbour in W . In this case all nearest neighbours are known to be within W and all necessary information for the index estimation of tree i is provided by the window. However, if the distance from the boundary is shorter than that from the n th neighbour, then it is possible that nearer trees to tree i occur outside W . In this case it is uncertain whether the window provides all the necessary information for index estimation, and therefore these trees are rejected in the analysis.

An indicator variable n_i is introduced, where $n_i = 1$ if tree i is to be accepted for analysis and $n_i = 0$ if it is to be rejected. The principal is illustrated in Fig. 3. Points P_1 and P_2 have the same distance $s_1 = s_2$ to the boundary of the window of observation W . Points P_{1n} and P_{2n} are the corresponding n th nearest neighbours in W . Because $c_1 > s_1$, $n_1 = 0$ and point P_1 is not accepted. Point P_2 , however, is accepted, since $c_2 < s_2$ and $n_2 = 1$.

As in the buffer method, there is an inevitable loss of a number of data points. To compensate for this loss when applying NN approaches, the accepted points are weighted to obtain an unbiased estimator. In statistical theory such weights are called Horvitz–Thompson weights (Horvitz and Thompson 1952), and the idea is to give rare events larger weights than the more common ones. The weights in this case are the inverses of the so-called reduced areas, F_i , of W . The reduced area, F_i , is a function of c_i and W . For the two most important cases it is

$$[2] \quad F_i = \begin{cases} (a - 2c_i)(b - 2c_i), & \text{rectangular plot} \\ (r - c_i)^2 \pi, & \text{circular plot} \end{cases}$$

where c_i is the distance between tree i and its n th nearest neighbour; a, b are the lengths of the sides of a rectangular sample plot; and r is the radius of a circular sample plot.

If c_i is small it is probable that $c_i < s_i$, F_i is large, and tree i receives only a small weight; for larger c_i , it is increasingly unlikely that $c_i < s_i$, F_i is small, and therefore larger weights are given. Theoretically, F_i can also be defined for windows of irregular shape.

In the estimation of M (eq. 1) by means of the NN1 estimator a further point process parameter plays an important role, the intensity or point density, that is, the mean number of points per unit area. Intensity is simply another way of

Table 1. Overview of individual tree indices used in this study, their formulae and the corresponding publications.

No.	Index (reference)	Biodiversity of	Formula	Variable definition
1	Uniform angle index (Hui and Gadaw 2002)	Tree positions	$W_i = \frac{1}{n} \sum_{j=1}^n w_j$	$w_j = \begin{cases} 1, \alpha_j < \alpha_0 \\ 0, \text{otherwise} \end{cases}$ α_j Angle pointing away from the reference tree i to neighbours j
2	Species mingling (Füldner 1995; Aguirre et al. 2003)	Tree species	$M_i = \frac{1}{n} \sum_{j=1}^n m_j$	$\alpha_0 = \frac{360^\circ}{n+1}$ $m_j = \begin{cases} 1, \text{species}_j \neq \text{species}_i \\ 0, \text{otherwise} \end{cases}$
3	DBH dominance (Gadow and Hui 2002; Aguirre et al. 2003)	Tree dimensions	$U_i = \frac{1}{n} \sum_{j=1}^n u_j$	$u_j = \begin{cases} 1, \text{DBH}_i > \text{DBH}_j \\ 0, \text{otherwise} \end{cases}$
4	DBH differentiation (Füldner 1995; Pommerening 2002)	Tree dimensions	$T1_i = 1 - \frac{\min(\text{DBH}_i, \text{DBH}_j)}{\max(\text{DBH}_i, \text{DBH}_j)}$	j is the first neighbour tree of i
5	Aggregation index (Clark and Evans 1954)	Tree positions	$R = \frac{\bar{r}_{\text{observed}}}{E(r)}$	$\bar{r}_{\text{observed}}$ Arithmetic mean of distances between reference trees and neighbours j
6	Coefficient of segregation (Pielou 1977)	Tree species	$S = 1 - \frac{N(b+c)}{ms + nr}$	$E(r) = \frac{1}{2\sqrt{\frac{N}{A}}}$ A Area of the observation window N No. of trees, $R \in [0, 2.1491]$ A, B Species N No. of trees m, n No. of reference trees r, s No. of times A and B serve as NN

Note: $n = 4$ is the number of neighbour trees used with indices 1–3. Values of indices 1–4 are distributed between zero and one.

See Table 2. $S \in [-1, 1]$

Fig. 2. Illustration of the mingling index for $n = 4$ neighbours. The neighbours are numbered according to increasing distance from the reference tree.

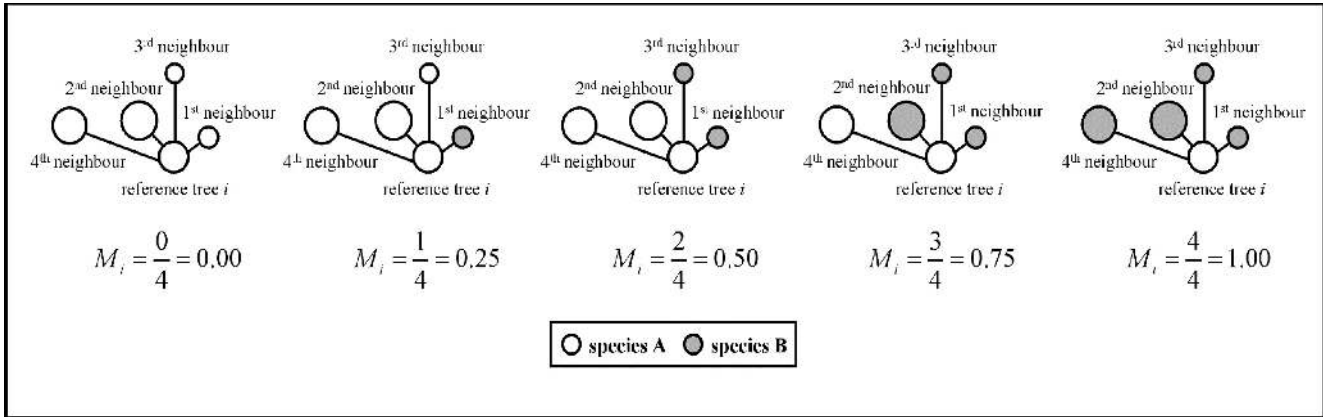


Table 2. A 2×2 table for the calculation of the coefficient of segregation (Pielou 1977).

Species of reference tree i	Species of nearest neighbour		
	A	B	
A	a	b	m
B	c	d	n
	r	s	N

expressing the forestry variable stems per hectare (SPH). It characterizes the density of trees in a certain area regardless of their marks or attributes such as their diameters.

The NN1 estimators of the mingling index class k and the mingling mean value are

$$[3] \quad \hat{m}_k = \frac{1}{\hat{\lambda}} \sum_{i=1}^N \frac{1_i(k)n_i}{F_i} = \frac{\overline{\lambda m_k}}{\hat{\lambda}}$$

and

$$[4] \quad \bar{M} = \frac{1}{\hat{\lambda}} \sum_{i=1}^N \frac{M_i n_i}{F_i} = \frac{\overline{\lambda M}}{\hat{\lambda}}$$

where

- $\hat{\lambda}$ estimated intensity of the point pattern (see eq. 6);
- N number of trees of the forest stand or monitoring plot (window of observation) under consideration;
- M_i mingling index of tree i ;
- \bar{M} estimator of the mean mingling index;
- F_i reduced area of the window of observation (see eq. 2);
- n_i indicator variable defining whether a tree's index is affected by plot edge (0) or not (1);
- $1_i(k)$ indicator function of mingling index class k assuming the value of one, if the neighbouring tree is of a different species, and zero otherwise.

It can be seen that M and m_k are not directly estimated but only $\overline{\lambda M}$ and $\overline{\lambda m_k}$, which are then divided by an intensity estimator $\hat{\lambda}$. Thus, the estimation is not unbiased but only ratio-unbiased.

The standard intensity estimator $\tilde{\lambda}$ is given by eq. 5.

$$[5] \quad \tilde{\lambda} = \sum_{i=1}^N \frac{N}{F}$$

where $\tilde{\lambda}$ is the number of points in W divided by the area F of W .

However, in this study a less precise estimator $\hat{\lambda}$ (eq. 6) was used. This is because $\hat{\lambda}$ has the same structure as $\overline{\lambda M}$ and it can be expected that fluctuations of $\overline{\lambda M}$ are evened out by fluctuations of $\hat{\lambda}$. This effect is weaker with $\tilde{\lambda}$:

$$[6] \quad \hat{\lambda} = \sum_{i=1}^N \frac{n_i}{F_i}$$

NN2 is closely related to NN1, having the advantage of perhaps being intuitively more appealing and easier to compute. However, it is not ratio-unbiased.

In this estimator the sum of all M_i of the accepted points (those with $n_i = 1$) is simply divided by the number of accepted points:

$$[7] \quad \tilde{m}_k = \frac{\sum_{i=1}^N 1_i(k)n_i}{\sum_{i=1}^N n_i}$$

and

$$[8] \quad \tilde{M} = \frac{\sum_{i=1}^N M_i n_i}{\sum_{i=1}^N n_i}$$

Experimental design

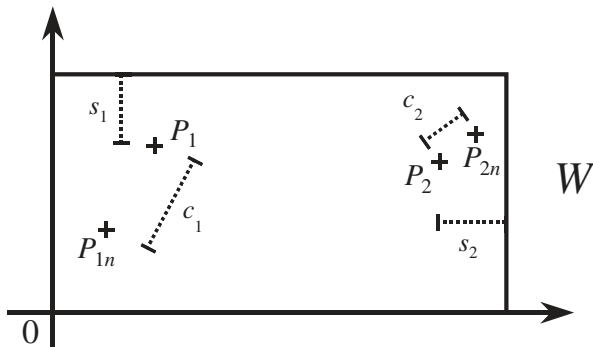
Edge-correction methods

Six different edge-correction methods were considered and compared in this study (Table 3). Estimation without edge-correction (method 0) was used as a control. In the control all trees within the monitoring plot or window of observation were processed as reference trees and their nearest neighbours were taken from within the window, ignoring the existence of outside trees. "Translation" or "periodic boundary conditions" (method 1) means that the plot W containing the tree locations is exactly replicated throughout space to

Table 3. Overview of the edge-bias compensation methods used in this study.

No.	Name	Reference	Comment
0	No edge-bias compensation	—	Control
1	Translation (= torus, periodic boundary conditions)	Torquato 2002; Radtke and Burkhart 1998	Fig. 1
2	Reflection	Pretzsch 2002; Radtke and Burkhart 1998	Fig. 1
3	Buffer zone	Diggle 2003; Gadow et al. 2003; Ripley 1981	5 m
4	NN1	Equations 3–4	—
5	NN2	Equations 7–8	—

Fig. 3. Illustration of the working principle of the indicator variable n_i . For point P_1 , $n_1 = 0$ (since its distance s_1 to the boundary of W is smaller than the distance c_1 to the nearest neighbour P_{1n} in the window), and for point P_2 , $n_2 = 1$.



form an infinite lattice (Torquato 2002). This method is the same as wrapping the window on a torus and joining opposite edges (see Fig. 1). All the trees in W are used as reference trees, and all their neighbours also come from W or its copies.

The reflection method (2) also replicates the pattern throughout space (see Fig. 1), but as the name implies, the additional locations of the trees are obtained by reflection along the plot edges. As in the translation method, all reference and neighbour trees are to be found in the window of observation or its reflected copies.

The buffer zone method (3) uses only the trees in the inner part of W as reference trees; the trees within a buffer strip of 5 m along the plot boundaries are used only in the calculation of the indices. NN1 and NN2 (methods 4 and 5) have already been explained in detail in the previous section. The last three edge corrections belong to the group of minus-sampling methods (Stoyan and Stoyan 1994; Stoyan et al. 1995).

Investigated point processes

Data from both point process models and real forests were used in the comparisons. Point process models guarantee knowledge of the true population values of the mean indices around which the sample values must necessarily fluctuate. Employing these models also enables unlimited repetitions and the statistical assumption of stationarity is guaranteed (Stoyan et al. 1995). To validate the results obtained from the point process models two real forest sites were included in the investigation.

Three different processes were simulated: a Poisson, a Matérn hard-core, and a Matérn cluster process (Stoyan et al. 1995). The second process is a process with some degree of regularity and a positive minimum interpoint distance.

The third is a classical cluster process, where the cluster centres form a Poisson process and the cluster points are randomly scattered in circles of constant radius around the centres. The three different processes and their specific parameters (see Table 4) were chosen to produce rather distinct patterns. The point process models approximate patterns found in real forest situations. In our study the Matérn cluster processes represent young forests of natural origin consisting of trees with heavy seeds or with heterogeneous soil conditions leading to a clustered arrangement of trees. Alternatively, this process may represent a failed re-stocking with species 1 and a moderate secondary invasion of species 2. The Matérn hard-core process can be thought of as representing older forests where it is impossible for trees to be located closer together than a certain minimum distance, perhaps because of inter- and intra-specific competition and (or) forestry thinning activities. The Poisson process is a situation between the young clustered and the more mature hard-core arrangement of tree positions. Such patterns, which are close to randomness, are common in forests (Tomppo 1986; Pommerening 2002).

For the simulation of all three processes, points were marked both dependently and independently with species and diameter attributes. Dependent marking leads to a correlation between tree locations and species and diameter attributes, while independent marking excludes such a correlation. Both cases can be found in real forests. The marks were assigned as follows:

- (i) Dependent marking: Within a circle of radius r_0 around each tree the number m of other trees was counted. If $m > 1$ the tree was assigned the species attribute 2, otherwise it was assigned species 1.
- (ii) Independent marking: The qualitative marks species 1 and species 2 were assigned randomly.

In both cases for the assignment of diameters, uniform random numbers were transformed to Weibull-distributed random numbers by applying the method developed by Nagel and Biging (1995). The species-specific scale and form parameters of the two-parameter Weibull function were estimated from a quadratic mean diameter by using linear relationships. In contrast to their original approach, the maximum diameter, which was required for estimating the form parameter, was estimated from the quadratic mean diameter, again by using a linear equation. The diameters assigned to each point are sampled from the corresponding species-specific Weibull functions by using the inverse transformation method. The Weibull distributions were restricted by a minimum diameter of 5 cm. The regression parameters used for the estimation of the parameters of the two-parameter Weibull function were taken from Nagel and Biging's (1995) original publication and are representative of northwest Ger-

many (where species 1 is oak (*Quercus robur* L., *Quercus petraea* (Matt.) Liebl.) and species 2 is beech (*Fagus sylvatica* L.)). Specific quadratic mean diameters were applied to each species (see Table 4) and as a result the diameter ranges of both species are also very different.

The purpose of the dependent marking procedure is to obtain isolated trees of species 1 and clusters of species 2. Such patterns appear, for example, in managed central European oak forests, where each dominant oak tree is surrounded by a group of smaller sized trees of one or more shade-tolerant species, for example, beech (Pommerening 2002).

Tables 4 and 5 summarize the specific processes used in the study. While the point patterns of Table 4 illustrate only 1 replication out of 10 000, the numerical values in the first six columns of Table 5 are the estimated true values, for the simulated models, of the indices for the processes involved. Formulae for these values could be derived mathematically in the case of Poisson processes with independent marks, but not for the more complicated processes. Thus, all values of Table 5 relating to the processes were determined numerically by simulation. For this purpose 100 very large forests (1000 m × 1000 m) were simulated to be sure that edge effects are negligible. The true index values were then determined as the arithmetic means of the 100 replications.

The mean values of the indices clearly reflect the qualitative differences between the processes. For example, for the aggregation index of Clark and Evans, a value larger than 1 for the Matérn hard-core processes indicates the given tendency towards regularity. For the Matérn cluster processes on the other hand the value is, as expected, smaller than 1.

The uniform angle index also behaves as expected, and in the case of the Poisson process a value of 0.5 is obtained exactly in accordance with its method of construction (Hui and Gadow 2002).

Real forest data

To validate the results of the statistically advantageous point processes, data from two monitoring sites in mixed oak–beech woodlands were used. Detailed information on the two monitoring plots is given in Fig. 4 and Table 5 (last two columns).

Manderscheid is a monitoring plot of 80 m × 80 m situated in the forest district of Manderscheid in the German federal state Rhineland–Palatinate (Pommerening 2002). The plot was surveyed in 1996 and consisted of a 118-year-old oak (*Quercus petraea*) stand intermingled with beech (*Fagus sylvatica*).

The second monitoring site, Boeselager, is also a mixed oak–beech stand of *Quercus robur* and *Fagus sylvatica*. The plot of 95 m × 116 m was also surveyed in 1996 and is situated in the Boeselager Estate in the Sauerland region (German federal state Northrhine–Westfalia; Bölsing 1996). At the time of surveying the oak trees were 151 years old and the beech trees 65 years old.

Manderscheid has a structure similar to the simulated model case Poisson with dependent marks, and Boeselager reflects Matérn hard-core with dependent marks.

Simulation procedure

The simulation for quantifying the performance of the six

different edge-correction methods (Table 3) was done in the following way:

- (i) Processes: Simulation of 10 000 forests (Poisson, Matérn hard-core, and Matérn cluster processes) in a window of 250 m × 250 m. Samples with smaller windows of decreasing size from 200 m × 200 m down to 30 m × 30 m were cut out of the centre.
- (ii) Real forest data: Simulation of 100 replications in the original plot windows. Samples with smaller windows of decreasing size from 70 m × 70 m (90 m × 90 m) down to 15 m × 15 m were cut, with a centre point randomly selected for each replication so that the sample window was located fully inside the original plot window.

Evaluation statistics

For the evaluation of the performance of the six edge-bias compensation methods (estimators) the bias and root mean squared error (RMSE) were used.

$$[9] \quad \text{Bias} = M - \hat{M}$$

$$[10] \quad \text{RMSE} = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (\hat{M}_k - M)^2}$$

where

M “theoretical” mean index value (i.e., “population” value), that is, “theoretical” mean species mingling;

\hat{M} index value, that is, species mingling, estimated from the 10 000 (100) replications;

\hat{M}_k index value, that is, species mingling, estimated from replication k ;

n number of replications; here $n = 10\,000$ (100).

Because of a lack of space the bias is displayed only for the uniform angle, Clark and Evans, and Pielou indices in the next section. Graphs showing the bias of the other indices can be obtained from the authors upon request.

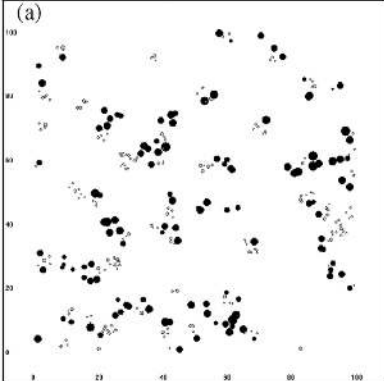
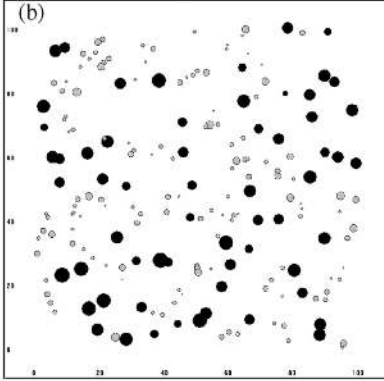
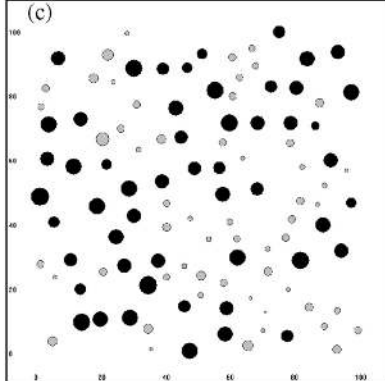
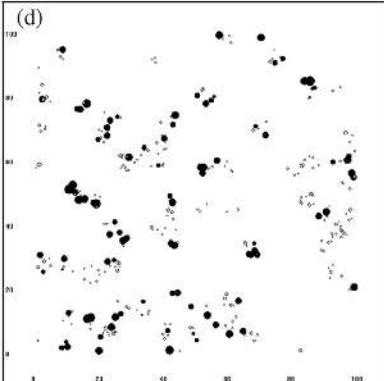
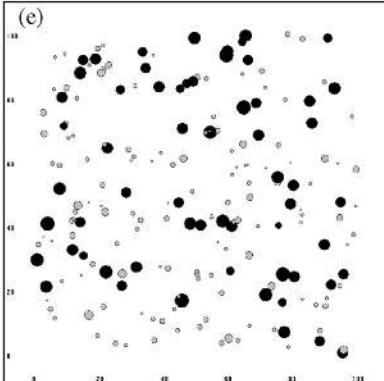
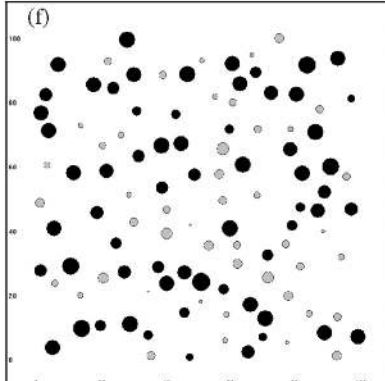
All results were obtained from the authors’ own software routines written in Pascal and Java and from the first author’s CRANCOD program.

Results and discussion

Figures A1–A12 (Appendix A) show the behaviour of the edge-correction methods expressed by bias (eq. 9) and RMSE (eq. 10) for each of the six indices depending on number of trees and point process model or monitoring plot. To enable comparisons, plot size was converted to number of trees using the known intensities. The Clark and Evans and the uniform angle index only account for the spatial arrangement of tree positions, so the results in Figs. A1, A2, A6, and A7 are the same for both dependent and independent marking. As expected, the RMSEs increase with decreasing numbers of trees.

The reaction of the six edge-correction methods of Table 3 to decreasing numbers of trees differs with the index involved. The speculative methods of no edge correction (0) and translation (1) (Pretzsch 2002; Diggle 2003) perform surprisingly well. This can be explained by the fact that these methods use all points in the window as reference points and accept some incorrect local values of the indices, which perhaps cancel each other out. Edge-bias investiga-

Table 4. Visualization (in a 100 m × 100 m window) of the processes involved in this study (species 1, black; species 2, gray).

	Matérn cluster	Poisson	Matérn hard-core
Parameters	$R = 3.0$ $\lambda = 0.03$ $dg_1 = 36$ cm $dg_2 = 13$ cm	$\lambda = 0.02$ $dg_1 = 60$ cm $dg_2 = 26$ cm	$h = 5.5$ $\lambda = 0.01$ $dg_1 = 75$ cm $dg_2 = 41$ cm
Illustrates	"Young forest"	"Poisson forest"	"Old forest"
Dependent	 $r_0 = 2.0$	 $r_0 = 5.7$	 $r_0 = 7.5$
Independent	 $p_1 = 0.270$	 $p_1 = 0.267$	 $p_1 = 0.542$

Note: "Dependent" and "independent" refer to the dependent and independent ways of marking tree locations with binary species and DBH. The specific parameters of the processes are given in a specific row and below the images. R , cluster diameter (m); λ , intensity; dg_1 , dg_2 : quadratic mean diameter (cm) of species 1 and 2, respectively; h , minimum distance between tree locations; r_0 , tree influence zone radius (m); p_1 , probability of species 1.

tions into competition indices came to similar conclusions (Monserud and Ek 1974; Radtke and Burkhart 1998). The translation method provides good estimators for all indices investigated. No edge correction is a good approach with most indices investigated and generally does not result in bad estimations. Two exceptions are the behaviour of the uniform angle and the Clark and Evans index. With these two indices, applying no edge correction is usually associated with a considerable bias throughout the range of numbers of trees observed. In the case of the uniform angle index, for points close to the window edge the nearest neighbours in the window are positioned towards the interior of the window, resulting in quite unusual angles, which spoil the statistical estimation.

The traditional method of reflection (2) leads, in many cases, to the highest RMSE (e.g., DBH dominance, DBH differentiation, Clark and Evans, and Pielou indices). The bias analysis reveals that the reflection method (2) results in

the largest bias values across all processes and real forest data, the only exception being the uniform angle index, for which reflection provides a good estimator.

Edge-correction methods that reject trees close to the boundaries of the observation window (methods 3–5) usually result in rapidly increasing bias values with decreasing numbers of trees, since the remaining number of trees becomes so small that ratio-unbiased estimators lose their good properties.

In most cases the buffer method (3) produces reasonable estimators (with the exception of cluster processes and Pielou index for hard-core processes and the mingling, DBH differentiation, Clark and Evans, and Pielou indices for the two real forest data sets). Where it performs well it often leads to similar RMSEs as those produced by the NN1 estimator.

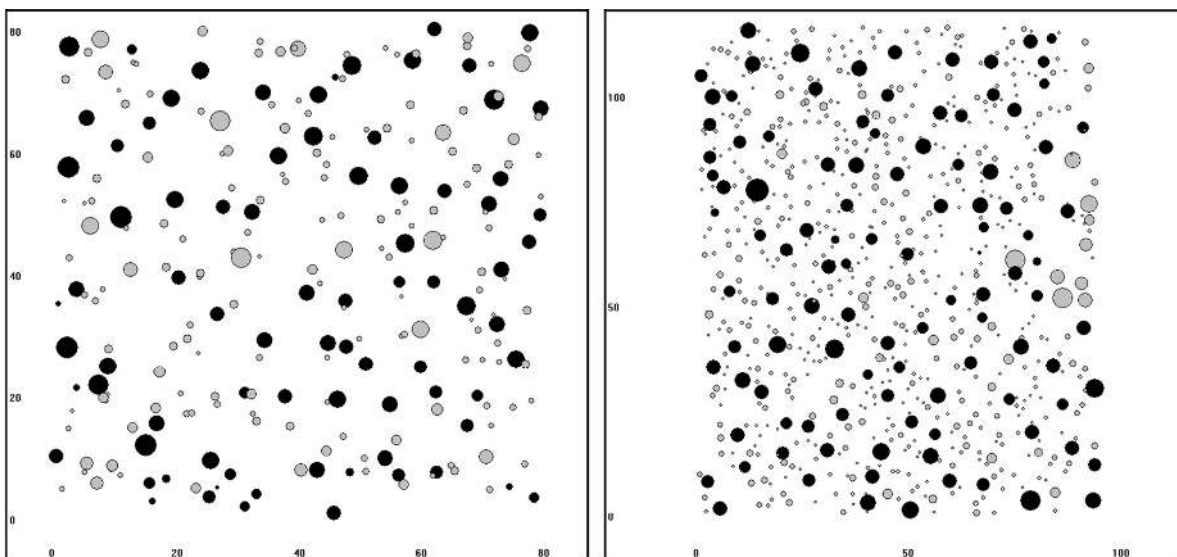
A comparison of the theoretically unbiased methods, NN1 (4) and NN2 (5), shows that throughout the application

Table 5. “Theoretical” arithmetic mean values (mean values of 100 very large forest stands were simulated in a 1000 m × 1000 m window) and population values (of the two real forest sites) of selected forestry characteristics.

Marking	Matérn cluster		Poisson		Matérn hard-core		Real forests	
	Dep.	Indep.	Dep.	Indep.	Dep.	Indep.	Manderscheid	Boeselager
UAI	0.60	0.60	0.50	0.50	0.42	0.42	0.47	0.46
Mingling	0.21	0.39	0.25	0.39	0.38	0.50	0.44	0.26
DBH dominance	0.53	0.50	0.53	0.50	0.52	0.50	0.53	0.51
T1	0.33	0.41	0.35	0.40	0.33	0.34	0.40	0.42
Clark and Evans	0.43	0.43	1.00	1.00	1.44	1.44	1.03	1.18
Pielou	0.44	0.00	0.38	0.00	0.22	0.00	-0.22	-0.15
BA (both species)	13.72	13.49	22.62	22.57	30.07	30.06	28.69	40.59
SPH (both species)	363.90	357.87	200.01	199.98	99.97	99.97	381.25	668.78
dg (both species)	21.33	21.91	37.94	37.91	61.89	61.88	30.95	27.80
BA (species 1)	9.97	9.80	14.97	14.94	24.12	24.11	17.12	23.48
SPH (species 1)	98.11	96.45	53.49	53.41	54.22	54.18	126.56	97.10
dg (species 1)	35.98	35.97	59.70	59.68	75.26	75.27	41.50	55.49
BA (species 2)	3.75	3.68	7.64	7.63	5.95	5.95	11.57	17.11
SPH (species 2)	265.79	261.42	146.52	146.57	45.75	45.79	254.69	571.69
dg (species 2)	13.40	13.40	25.77	25.75	40.71	40.69	24.05	19.52

Note: UAI, uniform angle index (eq. 1 in Table 1); mingling, species mingling (eq. 2 in Table 1); DBH dominance (eq. 3 in Table 1); T1, DBH differentiation (eq. 4 in Table 1); Clark and Evans, Clark and Evans (1954) aggregation index; Pielou, Pielou’s (1977) coefficient of species segregation; BA, basal area (m²); SPH, trees per hectare; dg, quadratic mean diameter (cm). “Dep.” and “Indep.” refer to the dependent and independent ways of marking tree locations with binary species and DBH. In the case of the real forest sites oak is coded as species 1 and beech as species 2.

Fig. 4. Visualization of the real forestry data used in this study (left: Manderscheid, right: Boeselager; oak (species 1), black; beech (species 2), gray).



range of this study both approaches behave similarly, although NN1 leads in a number of cases (uniform angle, DBH dominance, DBH differentiation, Clark and Evans, Pielou indices) to a smaller bias and can therefore be regarded as a safe option. Judging by the RMSE, NN1 provides a good estimator for the Clark and Evans index in particular. Both NN1 and NN2 perform well for DBH dominance, DBH differentiation, Clark and Evans, and Pielou indices if cluster processes are involved. The NN estimators (methods 4 and 5) omit many points close to the plot edge, which results in a large variability of the estimators. Their bias characteristics, however, especially those for NN1

(method 4), are very favourable except for a very small numbers of trees.

The bias analyses also clearly illustrate the difference between point patterns with dependent and independent marking. Figure A9, for example, shows that in the case of dependent marking there is a much larger differentiation between the different edge-correction methods, while the corresponding curves for the independent cases are very close together. Similar behaviour occurs with the mingling, DBH dominance, and DBH differentiation indices. The stronger the dependencies between tree locations and tree attributes the larger the bias differences of the estimators.

Especially when comparing DBH dominance and DBH differentiation indices it is evident that there are similar error patterns with indices quantifying equivalent aspects of spatial diversity. That the DBH of trees, especially, is correlated with the pattern of tree positions is well known from studies investigating competition indices (e.g., Pretzsch 2002). The error patterns of the indices accounting for species diversity and the diversity of tree positions are also similar but not as much as in the case of dimension diversity.

Conclusions

The optimal choice of edge-correction method is dependent on the algorithm of the index and on the spatial pattern of tree positions. This is particularly important when there are dependencies between the pattern of tree positions and the other tree attributes (marks), as this is often the result of competition processes and (or) woodland management. This highlights the importance of conducting sensitivity studies whenever new indices are developed or existing ones are modified.

The results clearly demonstrate the poor performance of the reflection method (2). This approach to edge correction should therefore be avoided in any investigation.

All methods that reject some sample trees (minus sampling, 3–5) usually lead to large bias values when the number of trees is small and therefore they should only be applied to samples containing a sufficiently large number of trees (≥ 100). These methods emphasize that the number of trees available for the index estimation is an important performance criterion.

Using the mathematical–statistically motivated edge-corrected NN1 estimator (method 4) with most indices leads to small bias values and can therefore be regarded as a secure option with sufficiently large numbers of trees. However, the NN1 and NN2 estimators can in many cases lead to larger RMSEs than those produced by ignoring the effect of edge bias or translation. Future research in edge-correction methods should develop the NN1 and NN2 approach by minimizing the loss of data and reducing the variability of the estimators.

Indices accounting for tree positions and especially those based on angles require edge-correction methods, no matter how large the sample size is. Translation, buffer zone, and NN1 methods are suitable methods to reduce error.

Plots to monitor indices of spatial biodiversity by means of spatial indices should be sufficiently large to minimize edge effects. This is especially crucial with long-term monitoring in forestry where the number of trees steadily decreases during the rotation period. Where small plot sizes are unavoidable, for example, in forest inventories with small-sized circular sample plots designed for other purposes (Gignoux et al. 1999), it is advisable to include the neighbours immediately outside the plot in the survey. Stoyan and Stoyan (1994) and Stoyan et al. (1995) refer to this method as plus-sampling. Plus-sampling is unbiased but requires additional measurements.

Acknowledgements

The authors thank the Research Committee of the School of Agricultural and Forest Sciences, University of Wales,

Bangor, for the financial support for two workshops that have facilitated the research cooperation between Freiburg and Bangor. Stephen T. Murphy (Forestry Commission Wales, University of Wales, Bangor) and Fergus L. Sinclair (University of Wales, Bangor) made thorough and constructive comments on earlier drafts of this paper. Stimulating discussions with Professor Hubert Sterba (BOKU University of Natural Resources and Applied Life Sciences, Vienna, Austria) greatly enhanced the study. This work is also a result of the activities of the EFI Project Centre CONFOREST.

References

- Aguirre, O., Hui, G.Y., Gadov, K., and Jiménez, J. 2003. An analysis of spatial forest structure using neighbourhood-based variables. *For. Ecol. Manage.* **183**: 137–145.
- Barot, S., and Gignoux, J. 2003. Neighbourhood analysis in the savanna palm *Borassus aethiopum*: interplay of intraspecific competition and soil patchiness. *J. Veg. Sci.* **14**: 79–88.
- Bauer, S., Wyszomirski, T., Berger, U., Hildebrandt, H., and Grimm, V. 2004. Asymmetric competition as a natural outcome of neighbour interactions among plants: results from the field-of-neighbourhood modelling approach. *Plant Ecol.* **170**: 135–145.
- Bölsing, S. 1996. Strukturbeschreibung von vier Eichenbeständen unterschiedlichen Alters. [Structural analysis of four oak stands of different age.] Diploma thesis, University of Göttingen. [In German.]
- Clark, Ph.J., and 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*. 2nd ed. Arnold, London.
- Donnelly, K. 1978. Simulation to determine the variance and edge effect of total nearest neighbour distance. *In Simulation methods in archaeology. Edited by I.R. Hodder*. Cambridge University Press, London. pp. 91–95.
- Füldner, K. 1995. Strukturbeschreibung von Buchen-Edellaubholz-Mischwäldern. [Describing forest structures in mixed beech-ash-maple-sycamore stands.] Ph.D. thesis, University of Göttingen, Cuvillier Verlag Göttingen. [In German.]
- Gadow, K.v., and Hui, G. 2002. Characterizing forest spatial structure and diversity. *In Proceedings IUFRO International Workshop: Sustainable Forestry in Temperate Regions*, Lund, Sweden. *Edited by L. Bjoerk*. pp. 20–30.
- Gadow, K.v., Hui, G.Y., Chen, B.W., and Albert, M. 2003. Beziehungen zwischen Winkelmaß und Baumabständen. [Relationship between the uniform angle index and nearest neighbour distances.] *Forstwiss. Centralbl.* **122**: 127–137. [In German.]
- Gignoux, J., Duby, C., and Barot, S. 1999. Comparing the performances of Diggle's tests of spatial randomness for small samples with and without edge-effect correction: application to ecological data. *Biometrics*, **55**: 156–164.
- Hanisch, K.-H. 1984. Some remarks on estimators of the distribution function of nearest neighbour distance in stationary spatial point processes. *Mathematische Operationsforschung und Statistik-Statistics*, **15**: 409–412.
- Horvitz, D.G., and Thompson, D.J. 1952. A generalization of sampling without replacement from a finite universe. *J. Am. Stat. Assoc.* **47**: 663–685.
- Hui, G.Y., and Gadov, K.v. 2002. Das Winkelmaß. Herleitung des optimalen Standardwinkels [The optimal standard angle of the

- uniform angle index.]. *Allg. Forst Holzwirtschaft. Ztg.* **173**: 173–177. [In German.]
- Martin, G.L., Ek, A.R., and Monserud, R.A. 1977. Control of plot edge bias in forest stand growth simulation models. *Can. J. For. Res.* **7**: 100–105.
- Monserud, R.A., and Ek, A.R. 1974. Plot edge bias in forest stand growth simulation models. *Can. J. For. Res.* **4**: 419–423.
- Nagel, J., and Biging, G.S. 1995. Schätzung der Parameter der Weibullfunktion zur Generierung von Durchmesserverteilungen [Estimation of the parameters of the Weibull function for generating diameter distributions.]. *Allg. Forst Holzwirtschaft. Ztg.* **166**: 185–189. [In German.]
- Neumann, M., and Starlinger, F. 2001. The significance of different indices for stand structure and diversity in forests. *For. Ecol. Manage.* **145**: 91–106.
- Penttinen, A., Stoyan, D., and Henttonen H.M. 1992. Marked point process in forest statistics. *For. Sci.* **38**: 806–824.
- Pielou, E.C. 1977. *Mathematical ecology*. John Wiley & Sons, New York.
- Pommerening, A. 2002. Approaches to quantifying forest structures. *Forestry*, **75**: 305–324.
- Pretzsch, H. 1998. Structural diversity as a result of silvicultural operations. *Lesnictvi*, **10**: 429–439.
- Pretzsch, H. 2002. *Grundlagen der Waldwachstumsforschung*. [Basics of forest growth science.] Parey Buchverlag, Berlin. [In German.]
- Radtke, P.J., and Burkhart, H.W. 1998. A comparison of methods for edge-bias compensation. *Can. J. For. Res.* **28**: 942–945.
- Ripley, B.D. 1981. *Spatial statistics*. John Wiley & Sons, New York.
- Stoyan, D. 2005. On estimators of the nearest neighbour distance distribution function for stationary point processes. *Metrika*. In press.
- Stoyan, D., and Stoyan, S. 1994. *Fractals, random shapes and point fields. Methods of geometrical statistics*. John Wiley & Sons, Chichester.
- Stoyan, D., Kendall, W.S., and Mecke, J. 1987. *Stochastic geometry and its applications*. John Wiley & Sons, Chichester.
- Tomppo, E. 1986. Models and methods for analysing spatial patterns of trees. *Commun. Inst. For. Fenn.* **138**.
- Torquato, S. 2002. *Random heterogeneous materials. Microstructure and macroscopic properties*. Interdisciplinary Applied Mathematics 16. Springer-Verlag, New York.
- Upton, G.J.G., and Fingleton, B. 1985. *Spatial data analysis by example. Volume I: Point pattern and quantitative data*. John Wiley & Sons, New York.
- Upton, G.J.G., and Fingleton, B. 1989. *Spatial data analysis by example. Volume II: Categorical and directional data*. John Wiley & Sons, New York.
- Windhager, M. 1997. Die Berechnung des Ek und Monserud (1974) Konkurrenzindex für Randbäume nach unterschiedlichen Berechnungsmethoden. [Calculation of the Ek and Monserud (1974) competition index for edge trees using alternative computing methods.] *In* Deutscher Verband Forstlicher Versuchsanstalten, Sektion Ertragskunde, Proceedings of the Jahrestagung 1997, Freiburg, Germany. *Edited by G. Kenk*. pp. 74–86. [In German.]

Appendix A

Appendix appears on the following page.

Fig. A1. Uniform angle index: results of the RMSE performance of the edge-bias compensation methods 0–5 (see Table 3) in relation to the point process models.

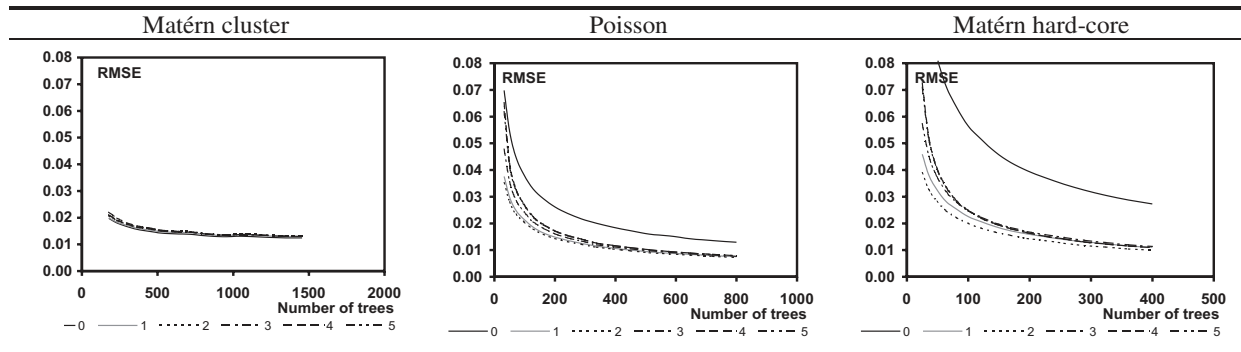


Fig. A2. Uniform angle index: results of the bias performance of the edge-bias compensation methods 0–5 (see Table 3) in relation to the point process models.

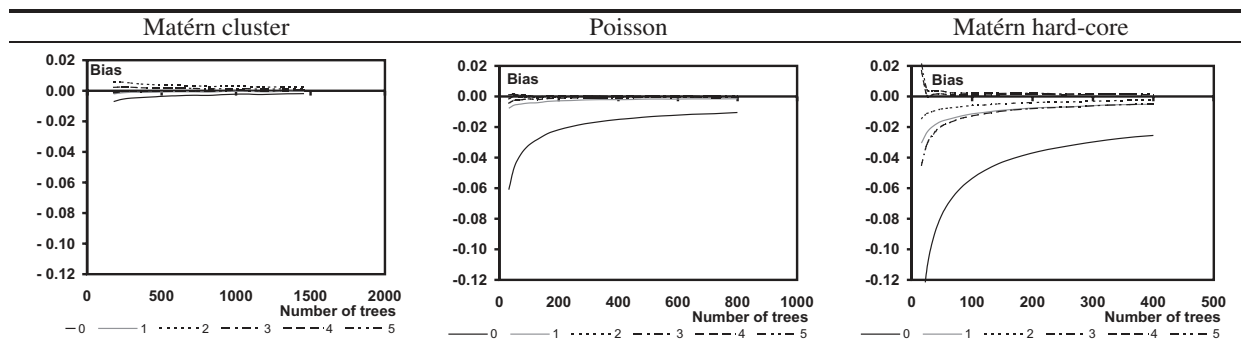


Fig. A3. Mingling index: results of the RMSE performance of the edge-bias compensation methods 0–5 (see Table 3) in relation to the point process models. “Dependent” and “independent” refers to the dependent and independent way of marking tree locations with binary species and DBH.

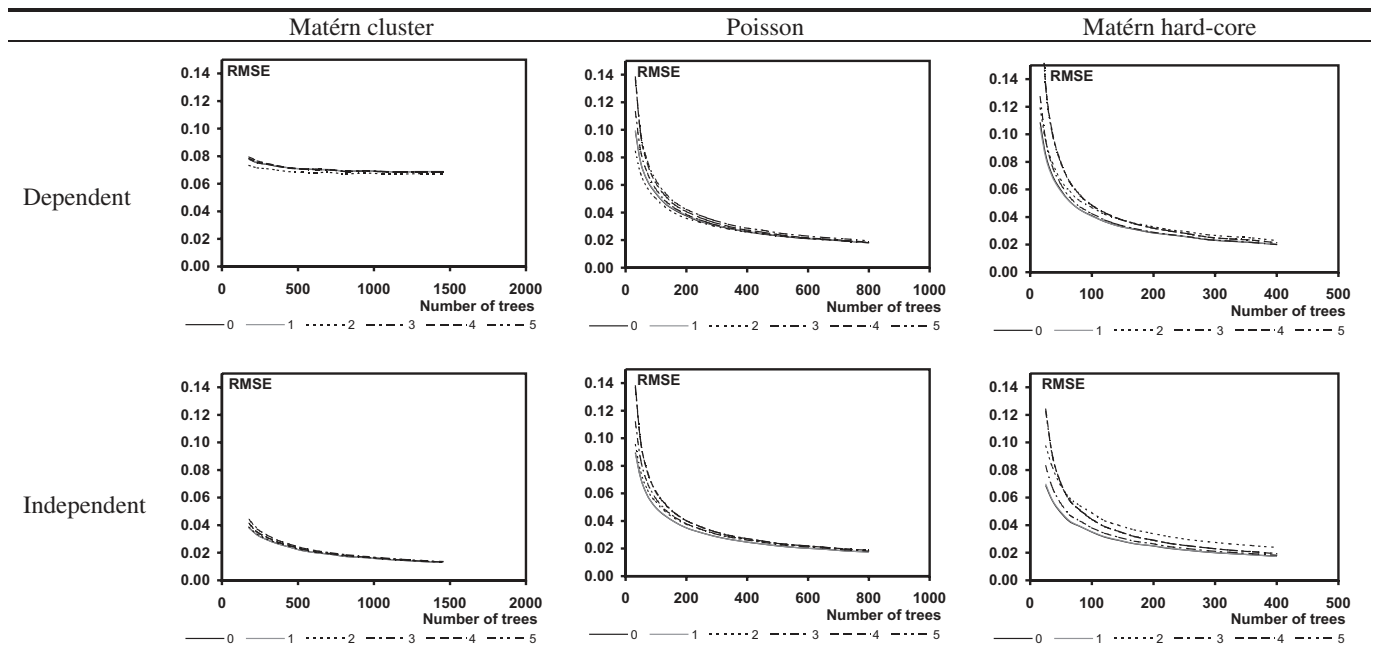


Fig. A4. DBH dominance index: results of the RMSE performance of the edge-bias compensation methods 0–5 (see Table 3) in relation to the point process models. “Dependent” and “independent” refers to the dependent and independent way of marking tree locations with binary species and DBH.

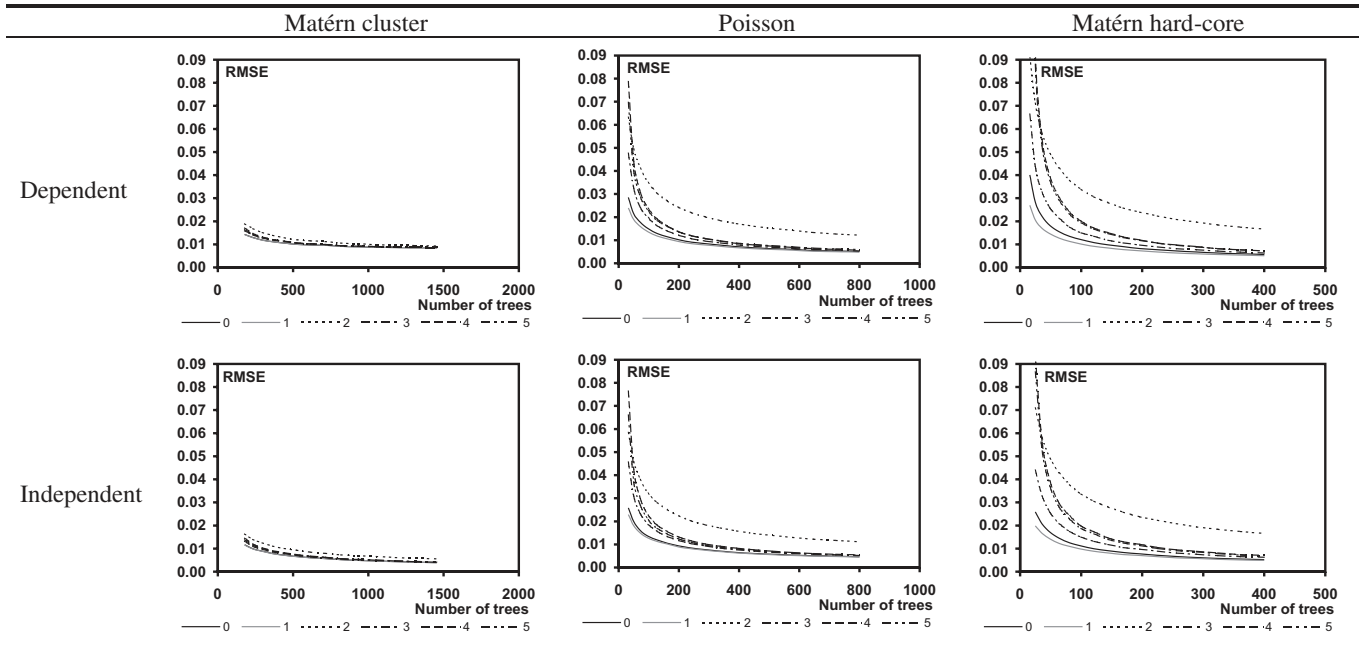


Fig. A5. DBH differentiation index: results of the RMSE performance of the edge-bias compensation methods 0–5 (see Table 3) in relation to the point process models. “Dependent” and “independent” refers to the dependent and independent way of marking tree locations with binary species and DBH.

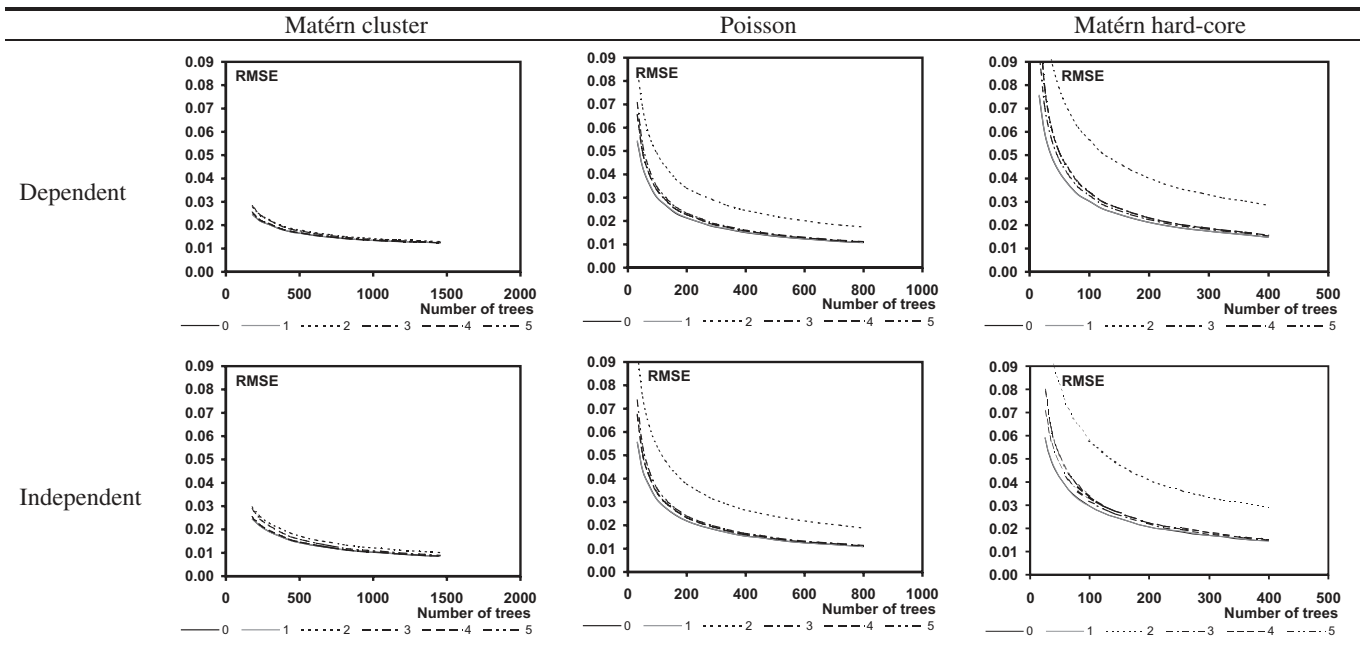


Fig. A6. Clark and Evans index: results of the RMSE performance of the edge-bias compensation methods 0–5 (see Table 3) in relation to the point process models.

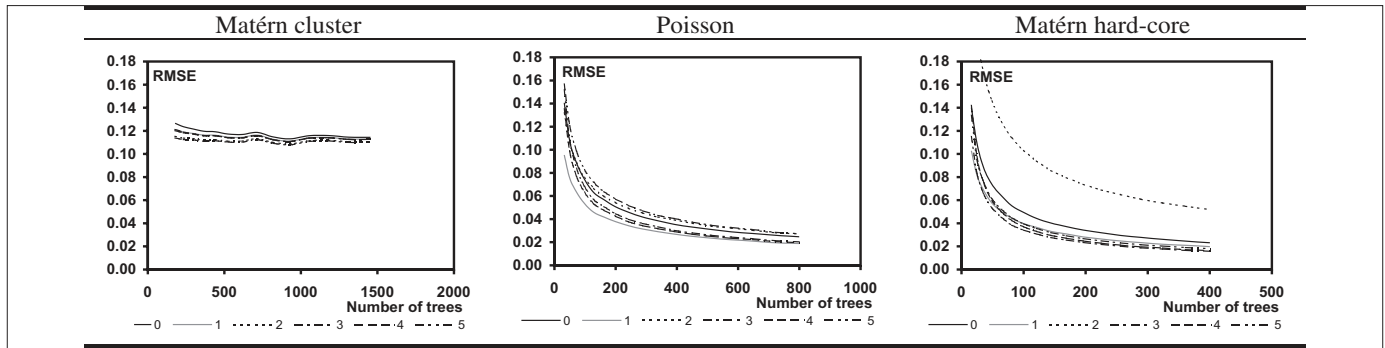


Fig. A7. Clark and Evans index: results of the bias performance of the edge-bias compensation methods 0–5 (see Table 3) in relation to the point process models.

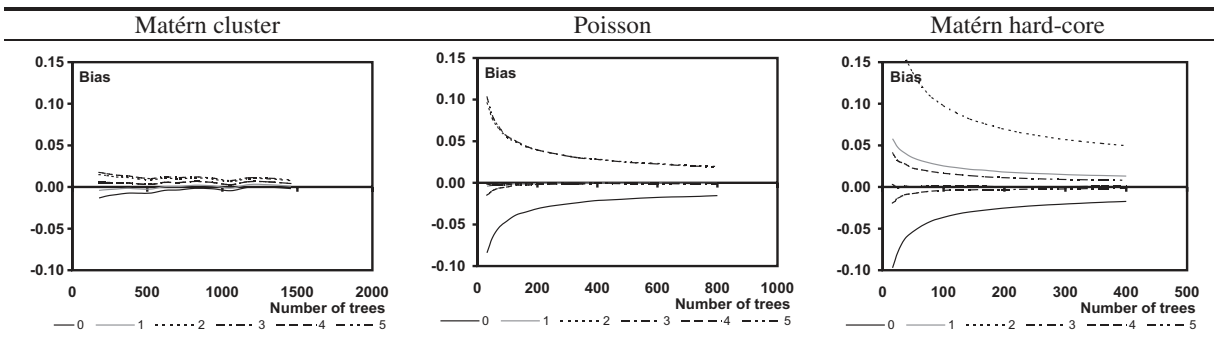


Fig. A8. Pielou index: results of the RMSE performance of the edge-bias compensation methods 0–5 (see Table 3) in relation to the point process models. “Dependent” and “independent” refers to the dependent and independent way of marking tree locations with binary species and DBH.

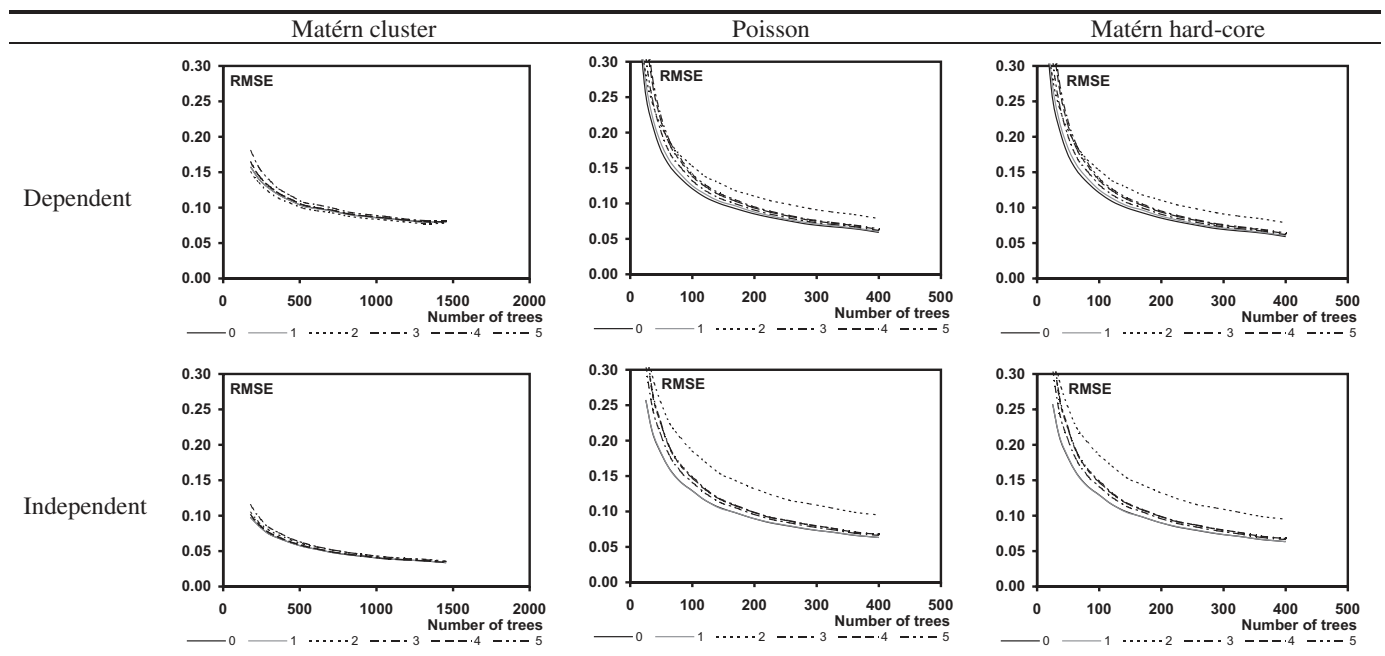


Fig. A9. Pielou index: results of the bias performance of the edge-bias compensation methods 0–5 (see Table 3) in relation to the point process models. “Dependent” and “independent” refers to the dependent and independent way of marking tree locations with binary species and DBH.

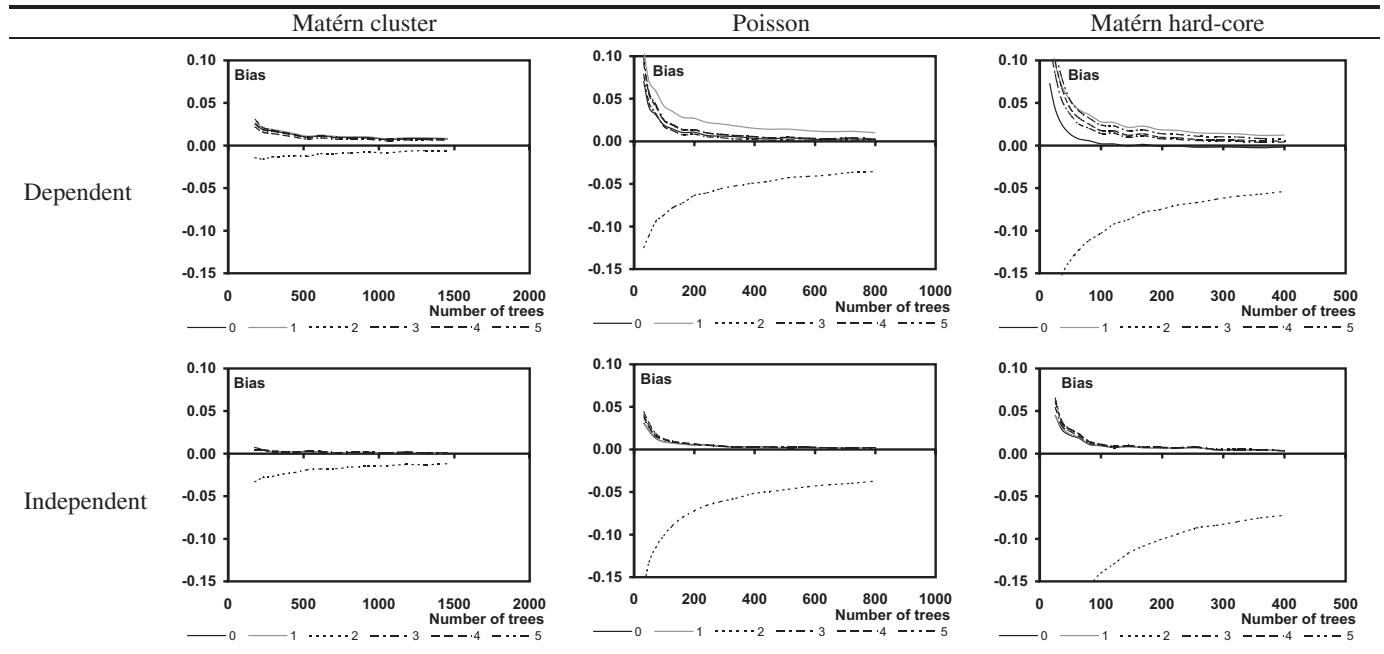


Fig. A10. Results of the RMSE performance of the edge-bias compensation methods 0–5 (see Table 3) and the uniform angle, mingling, and DBH dominance indices in relation to the two forestry data sets.

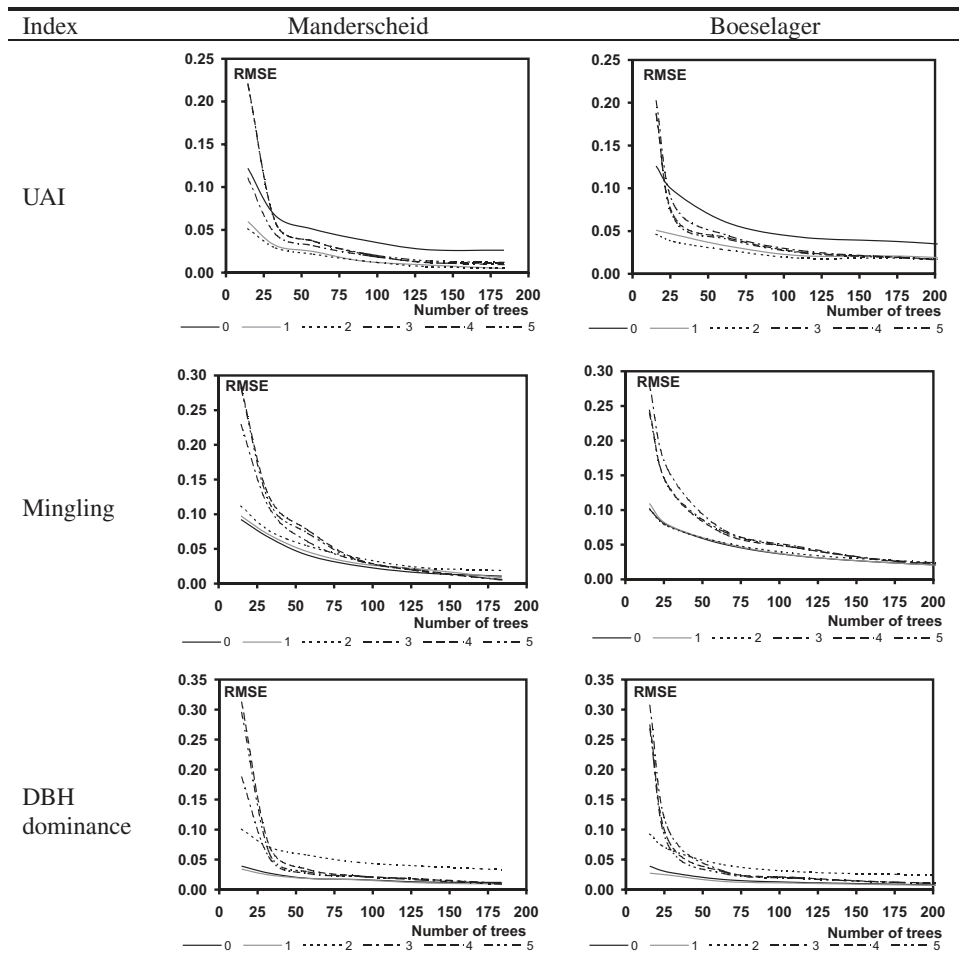


Fig. A11. Results of the RMSE performance of the edge-bias compensation methods 0–5 (see Table 3) and the DBH differentiation, Clark and Evans, and Pielou indices in relation to the two forestry data sets.

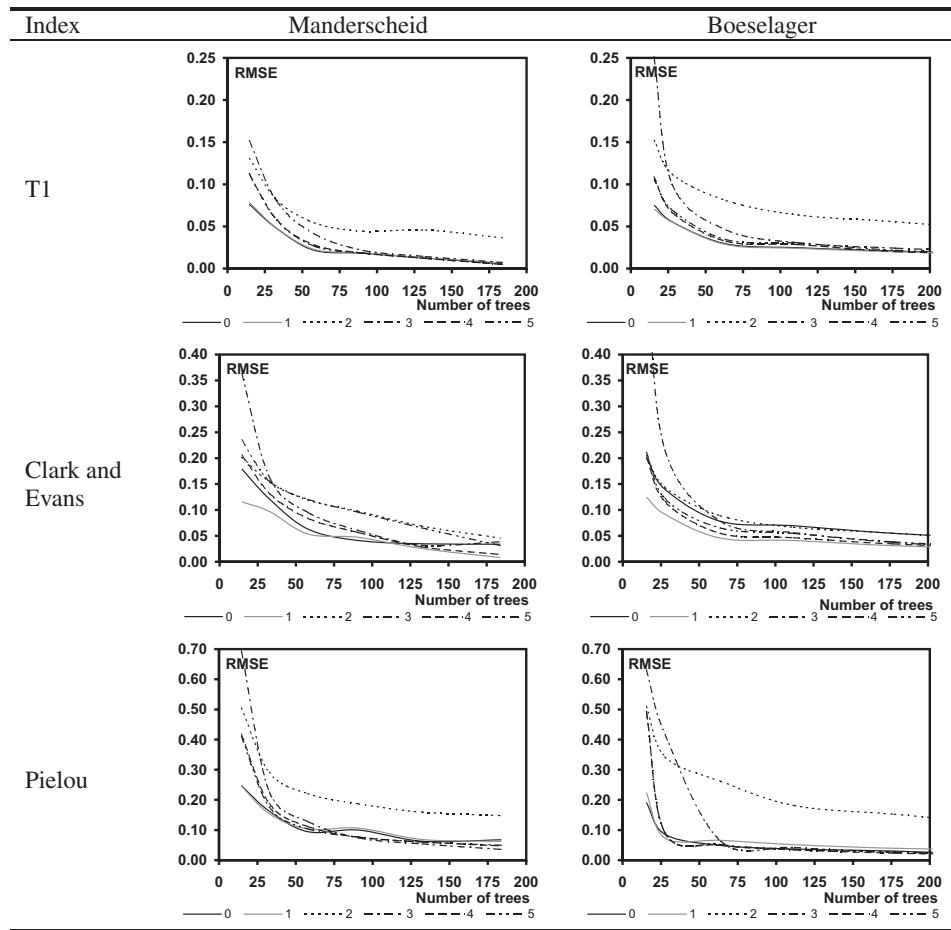


Fig. A12. Results of the bias performance of the edge-bias compensation methods 0–5 (see Table 3) and the uniform angle, Clark and Evans, and Pielou indices in relation to the two forestry data sets.

