A new nearest-neighbour index for monitoring spatial size diversity: The hyperbolic tangent index

Arne Pommerening[a], Janusz Szmyt[b], Gongqiao Zhang[c],*

[a] Swedish University of Agricultural Sciences SLU, Faculty of Forest Sciences, Department of Forest Ecology and Management, Skogsmarksgränd 17, Umeå SE-901 83, Sweden
[b] Poznań University of Life Sciences, Faculty of Forestry, Department of Silviculture, ul. Wojska Polskiego 71a, Poznań 60-625, Poland
[c] Research Institute of Forestry, Chinese Academy of Forestry, Key Laboratory of Tree Breeding and Cultivation of National Forestry and Grassland Administration, Box 1958, Beijing 100091, China

ARTICLE INFO

ABSTRACT

Understanding natural mechanisms of maintaining diversity is a crucial pre-requisite for successfully mitigating adverse effects of climate change such as the loss of diversity. To make such an understanding possible, both experiments and an effective, continued monitoring of diversity are required. Recently spatial measures of plant diversity have greatly contributed to the quality of diversity monitoring. In this article, we first reviewed existing principles of nearest-neighbour index construction and on this basis introduced a new spatially explicit size diversity index that is based on trigonometry, i.e. the hyberbolic tangent index. We discussed the index' mathematical reasoning by explaining its relationship to individual-based modelling and to other size diversity construction principles. Then we demonstrated the usefulness of the hyperbolic tangent index in indicating important interspecific relationships in mixed-species forest ecosystems. As part of studying the behaviour of the new size diversity construction principle we additionally found that there is a high correlation between the hyberbolic tangent index and absolute growth rates, i.e. the index is suitable both as a diversity and a competition index. Finally a detailed correlation analysis in a Norway spruce forest ecosystem with tree densities between 590 and 3800 trees per hectare made us understand that in most cases 7–10 neighbours are sufficient to consider when calculating the hyperbolic tangent index for explaining absolute growth rates. When using the index as an indicator of plant diversity only, smaller numbers of nearest neighbours may suffice. The index is straightforward to apply even, if the monitoring system used involves small circular sample plots.

1. Introduction

The analysis of plant species and size diversity offers great insight on the natural maintenance of biodiversity. The ecological insurance hypothesis, for example, states that biodiversity promotes greater insurance when communities allow for some functional redundancy of species in favourable environmental conditions (Yachi and Loreau, 1999; Matias et al., 2013) and is directly related to the provision of ecosystem goods and services as well as to ecosystem resilience (Perry et al., 2008). Often different species populations of the same ecosystem have distinctively different size ranges leading to species-size correlations (Wang et al., 2020; Pommerening et al., 2020). Much to our current concern, ongoing climate change is decreasing biodiversity on Earth at an unprecedented rate and effective monitoring of biodiversity is an essential pre-requisite for mitigating adverse effects (Krebs, 1999; Magurran, 2004).

Monitoring species diversity has a particularly long tradition (Gaston and Spicer, 2004) and the beginnings of measuring size diversity are much more recent (Ford, 1975; Weiner and Solbrig, 1984), where frequent semi-synonyms of size diversity include size inequality and size dominance. However, recent research has shown that spatial species and size diversity are much related and should be studied simultaneously (Pommerening and Uria-Diez, 2017; Wang et al., 2020; Pommerening et al., 2020). The uptake of point process statistics in quantifying spatial diversity has considerably improved our ability to identify spatio-temporal processes involved in the formation of plant diversity patterns and has contributed to a better understanding of the dynamics of plant diversity (Illian et al., 2008; Wiegand and Moloney, 2014; Pommerening and Grabarnik, 2019).

There are two main methods of constructing measures of spatial diversity, nearest neighbour (NN) and second-order methods. With both approaches, usually Euclidean distance is used to define spatial scales

---

* Corresponding author.
  E-mail address: zhanggongqiao@126.com (G. Zhang).

and relationships, but other distance definitions are also possible. As part of the nearest-neighbour approach, first a local index value is calculated for each plant based on the information derived from every plant itself and its neighbourhood. This index value describes the spatial diversity in the immediate vicinity of each individual, i.e. in its local neighbourhood (Rajala and Illian, 2012). Every plant within a given research plot acts once as subject or focus plant i. In this local index, the sizes of neighbouring plants need to be related to each other. Pommerening and Grabarnik (2019, p. 113) identified four principle methods, i.e. *size ratio, size comparison, size difference* and *size product* for relating the sizes of subject plants and nearest neighbours. The names of these methods are descriptors and give away the method: In most cases basic mathematical operations are used to relate the size of neighbouring plants to each other such as dividing, subtracting and multiplying. In the case of size comparison an indicator function is used to return a value of one for all cases where subject tree i is larger than neighbour tree j and zero otherwise. These scores are then summed over all neighbours j. In a second step, the local, plant-based indices are aggregated to produce summary characteristics for the whole plant community.

The objective of this work was to introduce and study a novel construction principle different from those used before in plant diversity monitoring that complements them and to apply this principle to a new nearest-neighbour index, i.e. the *hyperbolic tangent index*. The construction of this index is based on trigonometric principles. First we reviewed existing principles of quantifying size diversity and based on this brief review introduced the novel construction principle of the hyperbolic tangent index. Then we demonstrated the potential of the index to indicate different patterns of size dominance or size inequality in the context of mixed-species woodlands. Finally we analysed how the new hyperbolic tangent index relates to tree growth rates and how the correlation depends on the number of nearest neighbours.

2. Materials and methods

2.1. Existing measures of spatial size diversity

An early measure of spatial size inequality in plants was the size differentiation index introduced by Gadow (1993) as the mean of the ratio of smaller and larger plant sizes m of the k nearest neighbours subtracted from one (Eq. (1)). This index is based on the *size-ratio construction principle*.

$$T_i = 1 - \frac{1}{k} \sum_{j=1}^{k} \frac{\min(m_i, \ m_j)}{\max(m_i, \ m_j)} \tag{1}$$

Here m can be any quantifiable plant size measure, e.g. biomass, weight, height, stem diameter among others. The value of T_i increases with increasing average size difference between neighbouring trees. $T_i = 0$ implies that neighbouring trees have equal size.

Hui et al. (1998) and Aguirre et al. (2003) proposed an index using the *size-comparison construction principle*. This method turns the continuous size variables into a binary problem. The dominance index U_i (Eq. (2)) is defined as the mean fraction of plants among the k nearest neighbours of a given plant i that are smaller than plant i.

$$U_i = \frac{1}{k} \sum_{j=1}^{k} \mathbf{1}(m_i > m_j) \tag{2}$$

The indicator function returns 1, if the size mark of plant i exceeds that of neighbouring plant j, otherwise 0. U_i can have $k + 1$ possible discrete outcomes.

Based on the test function of the mark variogram Pommerening et al. (2011a) introduced the mark variogram index. This index is an example of the *size difference construction principle*:

$$V_i = \frac{1}{2k\sigma_m^2} \sum_{j=1}^{k} (m_i - m_j)^2 \tag{3}$$

Dividing by the size variance, σ_m^2, is a useful normalisation which eases the interpretation of index values and the comparison between different plants and plant communities. The smaller V_i the more similar the plant sizes considered are. In this case either both plants are small or both plants are large, no difference is made between these two scenarios.

Finally inspired by the test function of the mark correlation function, Davies and Pommerening (2008) introduced the size correlation index. Realising the *size-product construction principle* this index compares the mean product of the sizes of a subject plant and its k nearest neighbours with the squared arithmetic mean size, $\bar{m}^2$, of all plants in a given community (Eq. (4)).

$$C_i = \frac{m_i \sum_{j=1}^{k} m_j}{k \ \bar{m}^2} \tag{4}$$

Values larger than 1 indicate positive correlation which can be the result of similar sized plants at close proximity whilst negative correlation (indicated by $C_i < 1$) typically is the result of pairs of plants with large and very small sizes (e.g. a dominant subject plant surrounded by smaller plants) and pairs of plants with small sizes only.

2.2. A new measure of spatial size diversity based on trigonometric principles

In the context of individual-based models (IBM) using interaction-kernel functions for quantifying plant interaction, Schneider et al. (2006) and Adams et al. (2011) suggested a hyperbolic function accounting for the sharing of resources of two plants i and j:

$$f(m_i, m_j) = 1 + \tanh\left(\alpha \ \log_e \frac{m_j}{m_i}\right) \tag{5}$$

IBMs simulate populations and communities as being composed of discrete agents that represent individual organisms with sets of traits that vary among the agents following a strict bottom-up approach (Pommerening and Grabarnik, 2019, p. 217). In Eq. (5), m_i and m_j are again size variables of plant i and j. Plant i is the subject plant and plant j is a competitor or neighbour, i.e. the function describes the resource loss incurred by the subject plant. The main purpose of this function is to model the mode of interaction and parameter α introduces ecological symmetry ($\alpha = 0$) or asymmetry ($\alpha \to \infty$) (Freckleton and Watkinson, 2001). Function $f(m_i, m_j)$ increases the size dominance effect of plant j for $m_j > m_i$ and decreases size dominance of j otherwise. In a comparative analysis of different interaction functions carried out by Schneider et al. (2006), the inclusion of Eq. (5) led to a marked reduction in the summed root squared error. It was this use and the performance of $f(m_i, m_j)$ as a multiplier in spatial individual-based modelling (see also Pommerening and Grabarnik, 2019) that inspired the idea to study this function in greater detail and to explore how it would perform in the context of monitoring size diversity based on the nearest-neighbour principle.

Hyperbolic functions are transcendental functions and are closely related to exponential functions. The hyperbolic tangent function is denoted as $\tanh(x)$ and can be derived from hyperbolic sine and cosine functions as $\tanh(x) = \frac{\sinh(x)}{\cosh(x)}$. Another possibility to express the hyperbolic function is through the exponential function, i.e. $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1}$.

For expressing size dominance of subject plant i and for constraining the index value S_i to the interval [0, 1] it makes sense to swap indices i and j as in Eq. (6). We also discovered that this new variant of the hyperbolic tangent function can be expressed in terms of García's allotment function and also shares similarities with a function proposed earlier by Adler (García, 2014; Eq. (4) in Adler, 1996, see Appendix A).
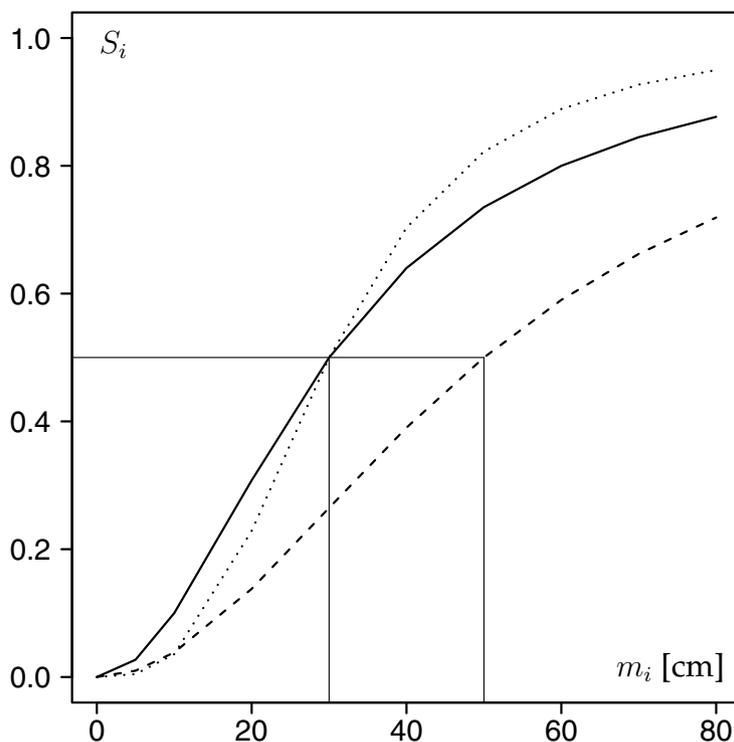
Fig. 1. Exploring the hyperbolic tangent index S_i with m_i ranging from 0 to 80 cm and $\alpha = 1.0$, $m_j = 30$ cm (solid curve), $\alpha = 1.0$, $m_j = 50$ cm (dashed curve). For the dotted curve $\alpha = 1.5$, $m_j = 30$ cm. The straight lines indicate $m_i = 30$ cm and $m_i = 50$ cm (after Pommerening and Grabarnik, 2019, p. 225).

$$S_i = \frac{1}{2k} \sum_{j=1}^{k} 1 + \tanh\left(\alpha \, \log_e \frac{m_i}{m_j}\right) = \frac{1}{k} \sum_{j=1}^{k} \frac{m_i^{2\alpha}}{m_i^{2\alpha} + m_j^{2\alpha}}$$

$$= \frac{1}{k} \sum_{j=1}^{k} \frac{1}{1 + \left(\frac{m_j}{m_i}\right)^{2\alpha}} \tag{6}$$

Factor 2 in the denominator of the first term before the sum symbol ensures that the index values lie between 0 and 1 (see Fig. 1). The second term is based on García's allotment function (García, 2014) and offers an easier interpretation of the hyperbolic tangent index. It shows that the index can be related to the size ratio principle, however, the ratio is more complex than the simple size ratio and involves power functions and sums. In the simplest case of $\alpha = 0.5$, the plant sizes are raised to the power of 1 and as a consequence the second and third term of Eq. (6) further simplify to $S_i = \frac{1}{k} \sum_{j=1}^{k} \frac{m_i}{m_i + m_j} = \frac{1}{k} \sum_{j=1}^{k} \frac{1}{1 + \frac{m_j}{m_i}}$.

However, we prefer retaining the variability of the mode parameter α, in order to be able to adjust the index to a wide range of different ecological situations. In the analyses presented in this paper, $\alpha = 1$ has proved to be a good assumption.

The larger the value calculated from Eq. (6) the larger the size dominance of plant i. The arithmetic mean is a naïve estimator of the population hyperbolic tangent index, but we recommend using the NN1 estimator instead (Eq. (7), Pommerening and Stoyan, 2006),

$$\hat{S} = \frac{1}{\hat{\lambda}} \sum_{i=1}^{N} \frac{S_i \times \mathbf{1}(dist_{ik} < c_i)}{F_i} \tag{7}$$

In Eq. (7), $dist_{ik}$ is the distance between plant i and its kth nearest neighbour, whilst c_i is the distance between plant i and the nearest point on the boundary of the observation window. The density estimator is calculated as $\hat{\lambda} = \sum_{i=1}^{N} \frac{\mathbf{1}(dist_{ik} < c_i)}{F_i}$ and F_i according to

$$F_i = \begin{cases} (a - 2 \, dist_{ik})(b - 2 \, dist_{ik}) & \text{for rectangular } W \\ (R - dist_{ik})^2 \, \pi & \text{for circular } W \end{cases} \tag{8}$$

In Eq. (8), a and b are the sides of a rectangular observation window and R is the radius of a circular observation window. For more detailed information probability density functions of S_i can be estimated based on the Epanechnikov kernel (see Illian et al., 2008; Pommerening and Grabarnik, 2019). The estimation of the probability density function follows the same principle as the estimator in Eq. (7).

As with any other spatial diversity index, the hyperbolic tangent index and its probability density function can be estimated for whole plant communities regardless of species, but also for individual species populations. This is achieved by subsetting the point pattern according to species and by estimating the index for each species subset separately.

There is yet another way of considering spatial species and size diversity at the same time, i.e. the interspecific case, where subject plant i is of one species and the neighbours j are of another or – in the case of many species – simply are of species different from that of subject plant i. This type of analysis is particularly useful when monitoring species interaction, because the interspecific analysis clearly highlights the spatial interactions between the most abundant or competing species in a given plant ecosystem. Since interspecific diversity analyses are not often published and they are particularly useful for exploring the properties of the trigonometric construction principle, we have applied them in this study.

In this paper, we used trees and tree stem diameter a 1.3 m above ground level as example plants and size variable, respectively. However, neither the trigonometric construction principle nor the hyperbolic tangent index is limited to plants with more or less single stems. Both can be applied to every plant with defined location co-ordinates such as the centre or the highest tip of a plant and quantifiable size attributes such as biomass or leaf area.

All calculations were carried out using our own R (version 3.5.1; R Development Core Team, 2018) and C++ code and the spatstat (Baddeley et al., 2016) package.

2.3. Data

Manderscheid is a management demonstration plot (80 × 80 m) situated in the West German federal state Rhineland-Palatinate (50.11 N, 6.8 E). There are two species in this plot, i.e. sessile oak (*Quercus petraea* MATT.) intermingled with beech (*Fagus sylvatica* L.). This forest stand has been mainly managed for quality oak, whilst beech was considered a minor by-product helping to improve the quality of oak timber (Pommerening, 2002; Pommerening and Uria-Diez, 2017). If left to natural devices, *F. sylvatica* would dominate the site. By supporting *Q. petraea* forest management also supports tree species diversity at the same time.

The *Södderich* data are from the Södderich, a part of the Göttinger Wald near the city of Göttingen in Germany (51.57 N, 10.08 E). The research plot (80 × 65 m) is dominated by beech (*Fagus sylvatica* L.), ash (*Fraxinus excelsior* L.) and sycamore (*Acer pseudoplatanus* L.), where for economic reasons the latter two species are favoured in forest management (Pommerening and Uria-Diez, 2017). Naturally *F. sylvatica* would also dominate this site and forest management therefore increases tree species diversity by removing *F. sylvatica* trees.

Norway spruce (*Picea abies* (L.) KARST.) spatio-temporal data in 16 plots (30 × 40 m each) at Karlstift in Austria (48.35 N, 14.46 E) with three re-measurement were studied using the relationship between hyperbolic tangent index and absolute growth rate (AGR). Originally the plots were part of a replicated thinning experiment. The trees have naturally regenerated and the plots are located at 930 m a.s.l. with a mean annual temperature of 4.5 °C and a mean annual precipitation of 950 mm. The plots were established in 1964 in predominantly even-aged *P. abies* and re-measured every five years until 2004 (Pommerening et al., 2011b).

All trees with a minimum breast-height diameter of 5 cm were included in the mapping at all three sites.

3. Results

3.1. Applications of the hyperbolic index for indicating species-size interactions

The probability density functions for *Q. petraea* and *F. sylvatica* in the Manderscheid woodland clearly showed the contrasting behaviour of the two species (Fig. 2A). Here we applied the hyperbolic tangent index in such a way that for all *Q. petraea* subject trees the nearest $k = 4$ neighbour trees were *F. sylvatica* and the neighbours of all *F. sylvatica* subject trees were in fact *Q. petraea* trees as described in Section 2.2. As can be concluded from the probability density functions in Fig. 2A, the population index values of this interspecific analysis were $\hat{S} = 0.79$ for *Q. petraea* and $\hat{S} = 0.22$ for *F. sylvatica*. Rejecting the null hypothesis of size mark independence (random labelling hypothesis; Pommerening and Grabarnik, 2019, p. 183f.), the corresponding p values were 0.006 and 0.004, respectively, i.e. both results are significant. Both the distributions and the population means indicate that *Q. petraea* trees dominate their *F. sylvatica* neighbours in terms of size and most likely also in terms of resource exploitation. In the context of the particular forest ecosystem at Manderscheid, this pattern stems from forest management where the economically more valuable and light demanding species *Q. petraea* is heavily promoted in thinnings on the expense of the economically less valuable, shade tolerant species *F. sylvatica* which would naturally dominate the site. In practical terms, competing *F. sylvatica* trees close to *Q. petraea* trees were selectively removed. *Q. petraea* would only regenerate naturally in such forest stands as a result of natural disturbances that interrupt the main forest canopy by creating gaps, however, *Q. petraea* seedlings and saplings would still face severe competition by *F. sylvatica*. Since large specimens of *Q. petraea* provide many more micro-habitats than those of *F. sylvatica*, this type of woodland management also promotes biodiversity in general. Still it is of interest to forest management here to maintain sub-dominant *F. sylvatica* trees (as opposed to a pure stand of *Q. petraea*), because the shade tolerant *F. sylvatica* services *Q. petraea* by suppressing epicormic growth along the stem axes that would otherwise de-value the oak timber.

A quite similar result is obtained from analysing the mixed-species Södderich woodland, where one species is again *F. sylvatica* that is analysed in relation to two other species, *F. excelsior* and *A. pseudoplatanus*, which are treated as one species group (Fig. 2B). Like in the Manderscheid woodland the *F. sylvatica* trees are those that are dominated by other species, in this case by *F. excelsior* and *A. pseudoplatanus*. This interpretation is confirmed by the population index values, which are $\hat{S} = 0.73$ for combined *F. excelsior* and *A. pseudoplatanus* and $\hat{S} = 0.32$ for *F. sylvatica*. The corresponding p values were 0.002 and 0.168, respectively. Apparently only the results for *F. excelsior* and *A. pseudoplatanus* are significant, i.e. here the null hypothesis of size mark independence (random labelling hypothesis; Pommerening and
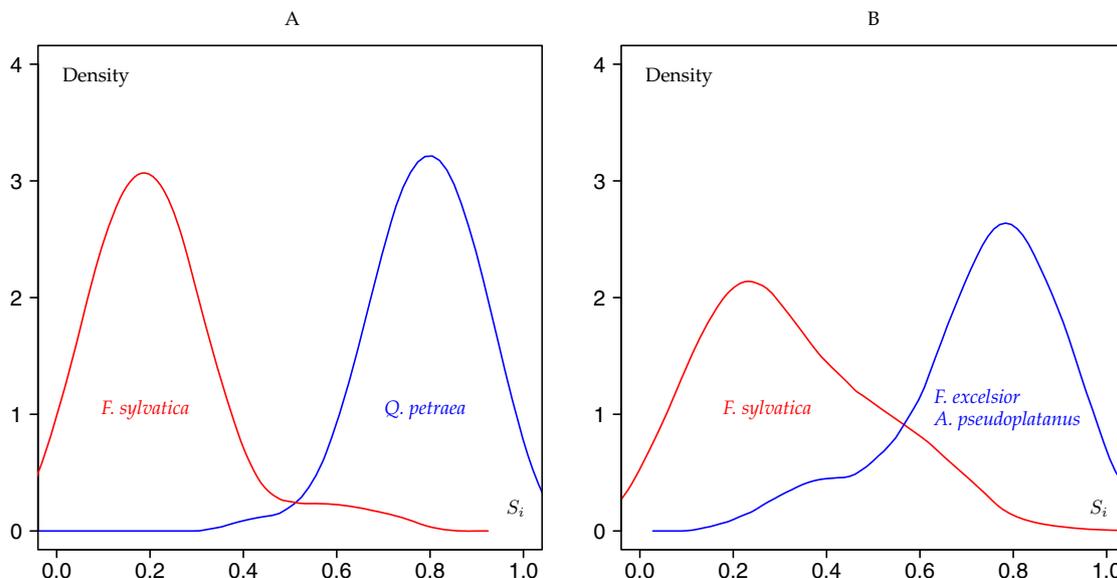
Fig. 2. The probability density functions of the hyperbolic tangent index S_i based on $k = 4$ neighbours, $\alpha = 1$ and NN1 edge correction (Pommerening and Stoyan, 2006) for the most abundant species at Manderscheid (A) and Södderich (B) woodlands.

Grabarnik, 2019, p. 183f.) can be rejected. The population means are closer together than in the case of Manderscheid and both S_i distributions are markedly more skewed in the Södderich woodland. This implies that the dominance separation between species is not as sharp as it is at Manderscheid. Also at Södderich the size dominance patterns are a result of selective near-natural forest management that favours *F. excelsior* and *A. pseudoplatanus* at the expense of *F. sylvatica* by removing any competitive *F. sylvatica* tree in proximity to *F. excelsior* and *A. pseudoplatanus* trees. The latter species would otherwise dominate and the former, more light demanding species would not be able to coexist with such large abundances.

3.2. Hyperbolic tangent index and growth rates

The hyperbolic tangent index S_i is correlated with both absolute and relative growth rates (AGR and RGR), which is not a common property of spatially explicit size diversity indices. We quantified AGR and RGR as mean annual growth rates as defined in Pommerening and Muszta (2016, Eq. (7) and Eq. (9)). For each plot we separately pooled the AGR and RGR data over two survey periods (1994 – 1999 and 1999 – 2004) as well as the values of the hyperbolic tangent index (measured in 1994 and 1999). In the data of the *P. abies* time series at Karlstift in Austria, we found a strong correlation between AGR and S_i with an asymptotic Pearson correlation coefficient between 0.5 and 0.9. There was also a linear relationship between RGR and S_i, however, in all 16 plots this was much weaker than the relationship with AGR. As an example we show the data for plot 42 in Fig. 3. In both cases the correlation is significant, but the correlation coefficient is much higher for AGR than for RGR.

When studying the correlation between hyperbolic tangent index and AGR, we were also interested in learning, how the Pearson correlation coefficient would change with increasing number of neighbours k. Again using the Karlstift *P. abies* time series data, we calculated the Pearson correlation coefficient for $k = 1, ..., 40$ neighbours. We deliberately used a very large upper number of k to be able to obtain reliable estimates of the upper asymptote of the relationship of interest. To minimise edge-effects periodic boundary conditions were applied in our simulations (Illian et al., 2008, p. 184; Pommerening and Grabarnik, 2019, p. 177).

For most of the 16 plots there was an increasing correlation trend with increasing number of neighbours (Figs. 4 and 5), particular for $k < 5$. The results also revealed an asymptotic behaviour of correlation with increasing k. In plots 11, 23, 33 and 41, the choice of k did not seem to matter much. In most plots, however, the choice of $k = 7, ..., 10$ appeared to be a reasonable compromise for avoiding low correlation coefficients due to small numbers of k.

In plots 11, 13, 33, 41 and 43, the calculated correlation coefficients reach the asymptote estimated from the saturation model (Michelis-Menten, 1913). Particularly in plots 20 and 21, the correlation coefficients remain much below the asymptote. Also the shapes of the model trend lines markedly differ much from plot to plot for reasons that are not easy to understand.

We compiled the base characteristics of all 16 plots in Table 1 to establish, whether any of them could potentially explain the correlation patterns. There was no characteristic, however, that particularly motivated the differences in the results of Figs. 4 and 5. Interestingly the highest asymptotic correlation coefficients were often achieved in plots with high densities, e.g. plots 10, 12, 20, 30 and 42. The mean aggregation index by Clark and Evans (1954) ranged between tree locations that were completely randomly dispersed (around 0.9 and 1.0) to very regular dispersal patterns (1.4 and 1.5) but did not help to explain the correlation patterns. Interestingly mean population hyperbolic tangent index $\hat{S}$ was always close to 0.50, which highlights that population values calculated regardless of species are not very informative. They are of greater value in mixed-species woodlands when calculated separately for each species (as in Section 3.1) and also the probability density distribution naturally offers more details than the simple population mean. It was also interesting to see that the mean *dbh* coefficient of variation generally was very high, indicating a high size diversity in all plots most likely as a result of planted *P. abies* trees mixing with *P. abies* trees that naturally colonised the plots but also as part of the quite different development stages captured in the data.

4. Discussion and conclusions

Our study has produced evidence that – based on the nearest neighbour principle – the trigonometric construction method indeed leads to a versatile spatial size diversity index. The construction principle and corresponding index complement and extend the existing provision if spatially explicit size diversity measures. Our theoretical

A   B
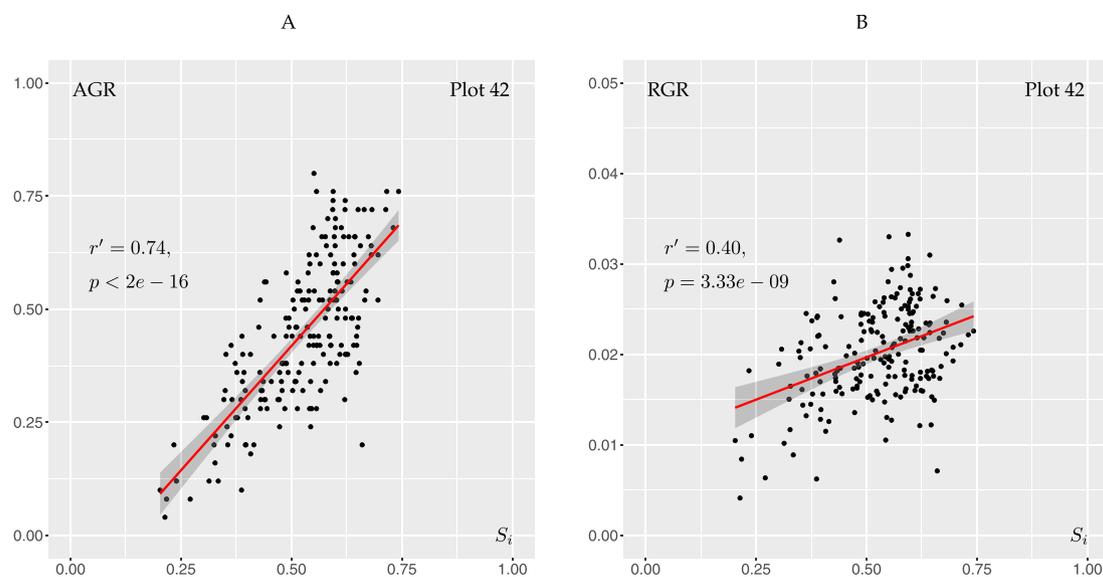
Fig. 3. Absolute (AGR, panel A) and relative (RGR, panel B) growth rates of the *P. abies* monitoring plot Karlstift (Austria) plotted over the hyperbolic tangent index S_i calculated based on $k = 7$ neighbours, $\alpha = 1$ and NN1 edge correction (Pommerening and Stoyan, 2006). The Pearson correlation coefficient r' and the associated p values are given in the graphs. The trendline was estimated through robust regression and the envelopes represent the standard error.

A

B
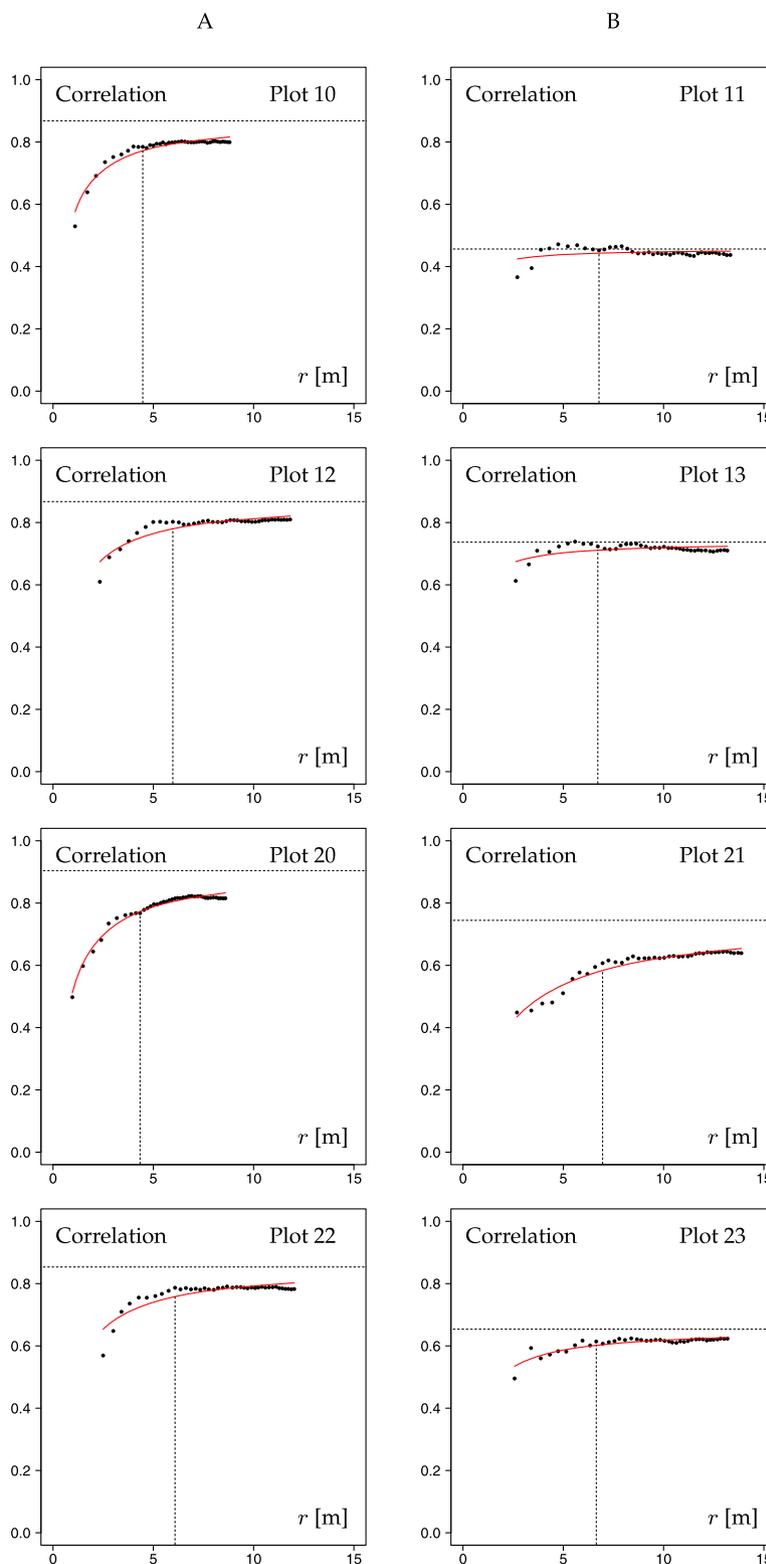
Fig. 4. Dependence of the correlation between S_i and the Pearson correlation coefficient on mean distance to the kth neighbour in the Karlsift monitoring plots 10-13 and 20-21. The dotted horizontal lines represent the asymptotic Pearson correlation coefficient as estimated by parameter a in the saturation model by Michaelis-Menten (1913), $r' = \frac{a\,r}{b+r}$ (shown as red trendline), where r' is the Pearson correlation coefficient and r the mean distance to the kth neighbour. The dotted vertical lines denote the correlation for the 10th nearest neighbour.

exposition in Sections 2.1, 2.2 and Appendix A have highlighted that the new index uses a construction principle previously unknown, i.e. the trigonometric construction principle. This principle leads to a real-valued index in contrast to the mark comparison construction principle.

In Appendix A, we have shown that this trigonometric construction principle is related to the size-ratio principle, however, the ratio used in S_i is more complex because of the powers and sums involved (Eq. (6), central term).
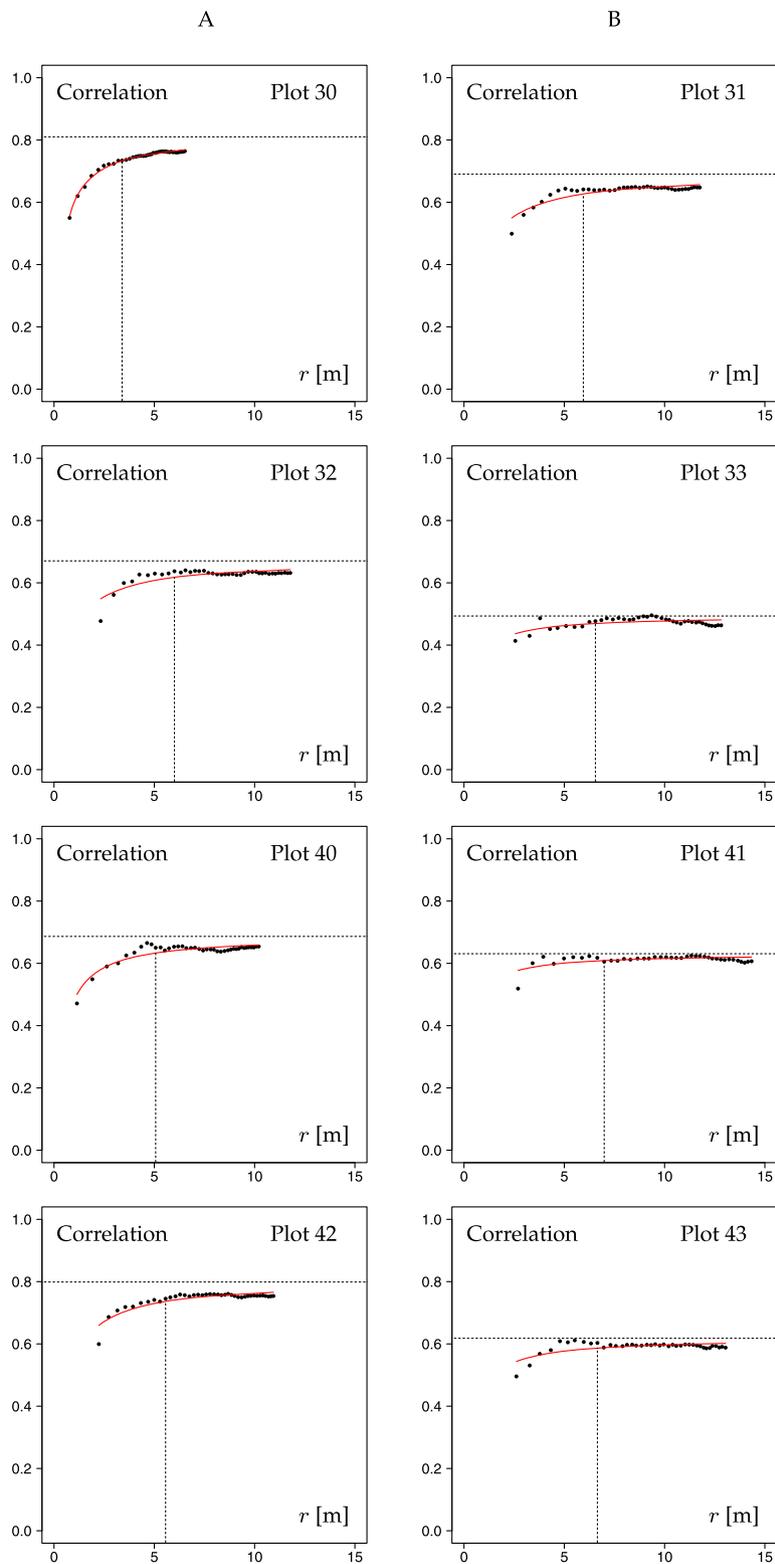
Fig. 5. Dependence of the correlation between S_i and the Pearson correlation coefficient on mean distance to the kth neighbour in the Karlsift monitoring plots 30-33 and 40-43. The dotted horizontal lines represent the asymptotic Pearson correlation coefficient as estimated by parameter a in the saturation model by Michaelis-Menten (1913), $r' = \frac{a\,r}{b + r}$ (shown as red trendline), where r' is the Pearson correlation coefficient and r the mean distance to the kth neighbour. The dotted vertical lines denote the correlation for the 10th nearest neighbour.

It is a particular strength of the hyperbolic tangent index to high-light size dominance differences between different species populations that occur in the same ecosystem. This is particular helpful when managing ecosystems, e.g. managing forest ecosystems for forestry purposes but also for conservation or recreation. A very useful appli-cation of this index is the monitoring of invasive or endangered species. For this purpose, it is particularly helpful to consider the probability density distribution as summary characteristic (Fig. 2). Population

Table 1

Basic characteristics of the *P. abies* time series data at Karlstift (Austria). The characteristics in columns 2–7 were averaged over the three surveys in 1994, 1999 and 2004. Clark and Evans refers to the aggregation index by Clark and Evans (1954) and $\hat{S}$ is the population hyperbolic tangent index as defined in Eq. (7). The asymptotic correlation coefficient is parameter a in the saturation model by Michaelis-Menten (1913), $r' = \frac{a\,r}{b+r}$, where r' is the Pearson correlation coefficient and r the mean distance to the kth neighbour.

Plot	Mean density [trees ha^{-1}]	Mean *dbh* [cm]	Mean *dbh* coeff. of variation	Mean basal area [m^2·ha^{-1}]	Mean $\hat{S}$	Mean Clark & Evans	Asymptotic correlation coefficient
10	2191.7	15.7	3.1	47.6	0.51	1.0	0.87
11	680.6	25.3	5.8	34.9	0.50	1.5	0.46
12	866.7	21.7	4.7	33.2	0.50	1.4	0.87
13	694.4	25.2	5.0	35.7	0.51	1.5	0.74
20	2277.8	14.1	2.7	41.7	0.50	0.9	0.90
21	650.0	25.6	6.4	34.1	0.51	1.4	0.74
22	794.4	23.7	5.1	35.6	0.51	1.5	0.85
23	694.4	26.7	4.9	40.2	0.50	1.4	0.65
30	3761.1	12.0	2.6	49.7	0.50	0.9	0.81
31	891.7	22.8	4.0	37.9	0.50	1.5	0.69
32	816.7	24.4	4.9	38.6	0.51	1.4	0.67
33	719.4	25.9	6.2	38.4	0.50	1.4	0.49
40	1602.8	17.3	2.7	45.1	0.50	0.9	0.69
41	591.7	27.9	5.9	37.1	0.50	1.4	0.63
42	969.4	21.7	5.0	36.0	0.51	1.5	0.80
43	697.2	25.6	5.1	37.0	0.50	1.5	0.62

characteristics summarising the size dominance situation in a single number such as $\hat{S}$ are only useful, if computed separately for several species in a mixed-species community. In this context, small numbers of nearest neighbours, e.g. $k = 4$ are sufficient. The application of the hyperbolic tangent index is therefore likely to be a good indicator of spatial species and size diversity relationships (Pommerening and Uria-Diez, 2017; Pommerening et al., 2020; Wang et al., 2020).

The hyperbolic tangent index can, however, additionally be used as a classic competition index with a view to explain and estimate relative and absolute growth rates. The relationships involved are linear and particularly using AGR as response variable can potentially lead to surprisingly large correlation coefficients. This interesting outcome tells us that the hyperbolic tangent index is in fact a good descriptor of competition, although the hyperbolic tangent function (Eq. (5)) was originally designed to model the mode of plant interaction complementing a kernel function that was supposed to quantify the strength of competition (Schneider et al., 2006). This implies that in future IBM applications one has to be careful when combining hyperbolic tangent and kernel functions, as both characteristics can compete with each other.

From a detailed correlation analysis we learned that the Pearson correlation index often increases with increasing number of nearest neighbours k used for computing S_i. Applying $k = 7, \dots, 10$ neighbours should lead to reasonable correlations in forest ecosystems with a similar range of densities (between 590 and 3800 trees per hectare). We also tested an approach of spatially balanced k akin to the concept of spatially balanced sampling (Stevens and Olsen, 2004; Grafström et al., 2012). In analogy to their work we optimised k individually for each tree i so that the distance to the kth nearest neighbour was similar for all trees and approached the mean distances to the $k = 1, \dots, 40$ nearest neighbours. This strategy made k dependent on local density so that k was large for trees situated at high local densities and small for trees in low local-density situations. Although this method successfully decreased the variance of distances to the k nearest neighbours, unfortunately the correlation coefficients did not improve compared to the results obtained from using fixed k instead (results not shown). This negative outcome confirmed that with a wide range of tree densities (between 590 and 3800 trees per hectare) it is indeed sufficient to apply $k = 7, \dots, 10$ neighbours as *fixed k* for all trees when calculating the hyperbolic tangent index for explaining absolute growth rates. Apparently the hyperbolic tangent index is suitable for basic AGR estimations based on simple nearest-neighbour information.

This study has defined a new versatile and meaningful construction principle for spatial size diversity indices that should prove useful in many situations where spatial size diversity needs to be monitored alongside spatial species diversity. The hyperbolic tangent index can also be applied to small circular monitoring plots, however, in the spirit of plus-sampling it is recommended to include any off-plot nearest neighbours in the analyses as well (see Pommerening and Grabarnik, 2019, p. 175ff.) to avoid edge-bias effects rather than using the NN1 estimator in this case.

Author contributions

A.P. designed and tested the new spatial size diversity index, programmed the required R and C++ code and analysed the data. A.P., J.S. and G.Z. interpreted and discussed the results. All authors contributed to the text.

Declaration of Competing Interest

None.

Acknowledgements

Appendix A

In this appendix, we explain the relationship between Schneider's hyperbolic tangent function (Schneider et al., 2006; Adams et al., 2011) and García's allotment function (García, 2014). Schneider's original hyperbolic function can be written as

$$1 + \tanh(\alpha(\log_e m_j - \log_e m_i)) = 1 + \tanh\left(\alpha \log_e \frac{m_j^\alpha}{m_i^\alpha}\right) \tag{A1}$$

Based on the tanh definition $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1}$ we can conclude[1] that

---

[1] $1 + \tanh(x) = 1 + \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^x + e^{-x}}{e^x + e^{-x}} + \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^x + e^{-x} + e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^x + e^x}{e^x + e^{-x}}$

$= \frac{2e^x}{e^x + e^{-x}}$

$$1 + \tanh(x) = \frac{2e^x}{e^x + \frac{1}{e^x}}. \tag{A2}$$

Eq. (A2) allows us now to substitute the second term of Eq. (A1) by

$$\frac{2 \frac{m_j^\alpha}{m_i^\alpha}}{\frac{m_j^\alpha}{m_i^\alpha} + \frac{m_i^\alpha}{m_j^\alpha}}. \tag{A3}$$

Finally, multiplying numerator and denominator by $m_j^\alpha m_i^\alpha$ we obtain

$$\frac{2 \, m_j^{2\alpha}}{m_j^{2\alpha} + m_i^{2\alpha}}. \tag{A4}$$

References

Adams, T., Ackland, G., Marion, G., Edwards, C., 2011. Understanding plantation transformation using a size-structured spatial population model. For. Ecol. Manage. 261, 799–809.

Adler, F.R., 1996. A model of self-thinning through local competition. Proc. Natl. Acad. Sci. U.S.A. 93, 9980–9984.

Aguirre, O., Hui, G.Y., Gadow, K.v., Jiménez, J., 2003. An analysis of spatial forest structure using neighbourhood-based variables. For. Ecol. Manage. 183, 137–145.

Baddeley, A., Rubak, E., Turner, R., 2016. Spatial Point patterns. Methodology and Applications with R. CRC Press, Boca Raton.

Clark, Ph.J., Evans, F.C., 1954. Distance to nearest neighbour as a measure of spatial relationships in populations. Ecology 35, 445–453.

Davies, O., Pommerening, A., 2008. The contribution of structural indices to the modelling of Sitka spruce*(Picea sitchensis)* and birch (*Betula* spp.) crowns. For. Ecol. Manage. 256, 68–77.

Ford, E.D., 1975. Competition and stand structure in some even-aged plant monocultures. J. Ecol. 63, 311–333.

Feckleton, R.P., Watkinson, A.R., 2001. Asymmetric competition between plant species. Funct. Ecol. 15, 615–623.

Gaston, K.J., Spicer, J.I., 2004. Biodiversity. An Introduction. Blackwell Publishing, Oxford.

Gadow, K.v., 1993. Zur Bestandesbeschreibung in der Forsteinrichtung. [New variables for describing stands of trees.]. Forst und Holz 48, 602–606.

García, O., 2014. A generic approach to spatial individual-based modelling and simulation of plant communities. Math. Comput. Forest. Nat.-Res. Sci. 6, 36–47.

Grafström, A., Lundström, N.L.P., Schelin, L., 2012. Spatially balanced sampling through the pivotal method. Biometrics 68, 514–520.

Hui, G.Y., Albert, M., Gadow, K.v., 1998. Das Umgebungsmaß als Parameter zur Nachbildung von Bestandesstrukturen. [Diameter dominance as a parameter for simulating forest structure.]. Forstwissenschaftliches Centralblatt 117, 258–266.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns. John Wiley & Sons, Chichester.

Krebs, Ch.J., 1999. Ecological Methodology, second ed. Addison Wesley Longman, Inc. Menlo Park, California.

Magurran, A.E., 2004. Measuring Biological Diversity. Blackwell Publishing, Oxford.

Matias, M.G., Combe, M., Barbera, C., Mouquet, N., 2013. Ecological strategies shape the insurance potential of biodiversity. Front. Microbiol. 3, 432.

Michaelis, M., Menten, M.L., 1913. Die Kinetik der Invertinwirking. [The kinetics of invertase action.]. Biochem. Z. 49, 333–369.

Perry, D.A., Oren, R., Hart, S.C., 2008. Forest Ecosystems, second ed. The Johns Hopkins University Press, Baltimore, pp. 632.

Pommerening, A., 2002. Approaches to quantifying forest structures. Forestry 75, 305–324.

Pommerening, A., Stoyan, D., 2006. Edge-correction needs in estimating indices of spatial forest structure. Can. J. Forest Res. 36, 1723–1739.

Pommerening, A., Gonçalves, A.C., Rodríguez-Soalleiro, R., 2011a. Species mingling and diameter differentiation as second-order characteristics. Allgemeine Forst- und Jagd-Zeitung. [Ger. J. Forest Res.] 182, 115–129.

Pommerening, A., LeMay, V., Stoyan, D., 2011b. Model-based analysis of the influence of ecological processes on forest point pattern formation – a case study. Ecol. Modell. 222, 666–678.

Pommerening, A., Muszta, A., 2016. Relative plant growth revisited: towards a mathematical standardisation of separate approaches. Ecol. Modell. 320, 383–392.

Pommerening, A., Uria-Diez, J., 2017. Do large forest trees tend towards high species mingling? Ecol. Inform. 42, 139–147.

Pommerening, A., Grabarnik, P., 2019. Individual-Based Methods of Forest Ecology and Management. Springer, Cham.

Pommerening, A., Wang, H., Zhao, Z., 2020. Global woodland structure from local interactions: new nearest-neighbour functions for understanding the ontogenesis of global forest structure. For. Ecosyst. 7, 22.

Rajala, T., Illian, J., 2012. A family of spatial biodiversity measures based on graphs. Environ. Ecol. Stat. 19, 545–572.

Development Core Team, R., 2018. R: a language and environment for statistical computing. R Found. Stat. Comput. http://www.r-project.org.

Schneider, M.K., Law, R., Illian, J., 2006. Quantification of neighbourhood-dependent plant growth by Bayesian hierarchical modelling. J. Ecol. 94, 310–321.

Stevens, D.L., Olsen, A.R., 2004. Spatially balanced sampling of natural resources. J. Am. Stat. Assoc. 99, 262–278.

Wang, H., Zhao, Z., Myllymäki, M., Pommerening, A., 2020. Spatial size diversity in natural and planted forest ecosystems: revisiting and extending the concept of spatial size inequality. Ecol. Inform. 56, 101054.

Weiner, J., Solbrig, O.T., 1984. The meaning and measurement of size hierarchies in plant populations. Oecologia 61, 334–336.

Wiegand, T., Moloney, K.A., 2014. Handbook of Spatial Point Pattern Analysis in Ecology. CRC Press.

Yachi, S., Loreau, M., 1999. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. Proc. Nat. Acad. Sci. USA 96, 1463–1468.