

Characterising forest spatial structure through inhomogeneous second order characteristics

C. Comas · M. Palahí · T. Pukkala ·
J. Mateu

Published online: 18 March 2008
© Springer-Verlag 2008

Abstract Point process theory plays a fundamental role in the analysis and modelling of spatial forest patterns. For instance, the Ripley's K function and its density with respect to the area, i.e. the pair correlation function, have been extensively used to analyse and characterise stationary forest configurations. However, the stationarity condition is not often met in practice when analysing real data. Thus, the development and application of new statistics to measure the degree of inhomogeneity suggests the use of inhomogeneous statistics to describe forest stands. In this paper, we restrict our attention to the inhomogeneous pair correlation function in the context of replicated spatial data. We then analyse the spatial configuration of pure and mixed conifer stands in a case study in Central Catalonia, North-East of Spain. Our results suggest that whilst *P. sylvestris* tend to be aggregated for short inter-tree distances, *P. nigra* and *P. halepensis* keep a minimum inter-event distance between trees. Regarding the mixed stands, trees of distinct species tend to be segregated from each other. Tentative explanations for these results are related

with site properties, competition effects, shade tolerance and silviculture practices applied in this forest region.

Keywords Forest structure · Inhomogeneous pair correlation function · Non-stationary point processes · Replicated forest data · Ripley's K function

1 Introduction

Forests are dynamic biological systems that are continuously changing. Typically, forest management is based on information about current and future resource conditions. Whilst forest inventories provide statistical information about current resource conditions such as timber volume and forest composition, it is clearly necessary to use predictive modelling schemes in order to update inventory information, and hence understand, forest dynamics over time periods spanning several decades (Peng 2000). To do so, yield and growth models have been used extensively to project current forest conditions into future yields. The use of forest modelling has also proved to be an invaluable tool in the understanding of complex ecological forest dynamics.

In spite of the large number of growth-yield forest models developed during the last decades (see, for instance, Stage 1973; Wykoff et al. 1982; Teck et al. 1996; Monserud and Sterba 1996; Sterba and Monserud 1997; Palahí et al. 2003) few models have considered explicit spatial information in their basic formulation (Ek and Monserud 1974; Pacala et al. 1993; Pukkala et al. 1998). Distance-dependent tree-level models (i.e. trees are the basic unit of analysis) not only improve the predictive power of these formulations and permit to analyse spatially

C. Comas
Centre Tecnologic Forestal de Catalunya,
Pujada del Seminari s/n, 25280 Solsona, Spain

C. Comas · J. Mateu (✉)
Department of Mathematics, Universitat Jaume I,
Campus Riu Sec, 12071 Castellon, Spain
e-mail: mateu@mat.uji.es

M. Palahí
EFI Mediterranean Regional Office, Passeig Lluís Companys 23,
08010 Barcelona, Spain

T. Pukkala
ForEco Technologies, Avinguda Diagonal 416, Estudio 2,
08037 Barcelona, Spain

explicit silvicultural problems such as plantation and thinning strategies, but also allow the study of complex forest dynamics (Miina et al. 1991; Pukkala et al. 1998; Moustakas and Hristopoulos 2007; Renshaw et al. 2008). These models are also necessary to generate spatially explicit forest patterns (realistic synthetic data), to simulate and study realistic silvicultural operations such as thinning and regeneration strategies, and to compare inventory designs.

Typically, distance-dependent tree-level models require tree coordinates as well as individual qualitative and quantitative tree variables such as tree species, and tree height and trunk diameter, respectively, in order to simulate forest growth. As such, a “natural” way to describe and analyse forest dynamics is through the development of spatial (marked) point processes. A realisation of a spatial marked point process consists of a set of points x_i with an associated mark $m(x_i)$ in a bounded region A (Stoyan et al. 1995).

A first step before a distance-dependent forest model is confronted against real forest data is to obtain summary statistics to understand the underlying forest spatial configuration. Knowing the spatially explicit forest structure is clearly necessary if we are to define a realistic forest predictive scheme based on distance-dependent competition indices. A popular method to analyse and characterise homogeneous (i.e. stationary and/or isotropic) spatial forest patterns is the second order reduced moment measure, or also called the Ripley’s K function (Ripley 1976). Heuristically, this function defines the expected number of further points within a ball $b(0,r)$ centred at an arbitrary point 0 with radius r , providing information about the spatial point structure. This function has been extensively used to analyse forest spatial patterns (Ripley 1977; Péliissier 1998; Chen and Bradshaw 1999; Youngblood et al. 2004; Aldrich et al. 2003; Camarero et al. 2000), the spatial structure of forest products (Nanos et al. 2001) and the distribution and severity of infected trees (Shaw et al. 2005).

The use of the Ripley’s K function as defined by Ripley (1976), and also presented by popular ecological text books such as that of Fortin and Dale (2005), is restricted to stationary forest patterns. However, real forest configurations are seldom stationary. Soil fertility, the presence of a river or merely environmental heterogeneity can promote inhomogeneous forest structures. To analyse inhomogeneous point patterns based on non-parametric summaries, Baddeley et al. (2000) formulate an inhomogeneous counterpart version of the Ripley’s K function. However, this method has not been widely used to analyse forest patterns, probably due to the difficulty of applying this inhomogeneous K function. Also other statistical techniques have recently been considered to describe

inhomogeneous forest patterns. Explicit parametric point process models such as inhomogeneous Poisson and Gibbs process, and Cox processes have been applied to analyse inhomogeneous forest configurations (see Leps and Kindlman 1987; Stoyan and Stoyan 1998; Møller et al. 1998). However, two weak points in this context are the following: (a) extension of inhomogeneous characteristics to replicated spatial point patterns, and (b) testing for inhomogeneity in spatial patterns. To the best of the authors’ knowledge, no research in these two mentioned points has been published so far. The spatial heterogeneity has been dealt in several recent papers. For example, under the context of increasingly enriched spatiotemporal data, Li et al. (2007) suggest an information–fusion method to identify patterns of spatial heterogeneity. Onof et al. (2000) propose using inhomogeneous Poisson-cluster processes to model rainfall data.

We thus focus in this paper on extending inhomogeneous characteristics to replicated data, and testing for inhomogeneity. We illustrate the use of the inhomogeneous replicated K statistics for pure and mixed stands in a case study in the Pyrenees North-East of Spain. In particular, we restrict our attention to the Ripley’s K density, i.e. the pair correlation function (see Stoyan and Stoyan 1994). Thus, we illustrate the inhomogeneous pair correlation function (also defined by Baddeley et al. 2000) and present an edge-corrected estimator for the inhomogeneous partial pair correlation function for this case study. To justify the use of inhomogeneous statistics, we develop a simple statistic to measure the degree of inhomogeneity based on the difference between the edge-corrected estimator of the intensity function and the estimator of the intensity, under stationarity. In addition, given that this real forest data consists of several point patterns involving plots with one or two species, we analyse the resulting inhomogeneous statistics for a given region A as replicates of an underlying process. We thus briefly discuss the applicability of overall estimates for the Ripley’s K function (see also, Diggle 2003).

The main aim of the present paper is to illustrate inhomogeneous variability measures with real forest patterns and discuss the applicability of overall estimates of such statistics for replicated forest data. To do so, we present the main inhomogeneous variability statistics including the inhomogeneous Ripley’s K function, the inhomogeneous pair correlation and the inhomogeneous partial pair correlation functions, introduce a new statistic to measure the degree of inhomogeneity, and discuss and apply overall estimates for replicated forest data. Section 2 presents a brief theoretical setup of point process theory and its relation with forest statistics while introducing the main inhomogeneity measures. In Sect. 3, point processes’ statistics for replicated data are presented, and a measure to

characterise spatial forest inhomogeneity is developed in Sect. 4. Finally, Sect. 5 presents the statistical analysis of the case study using edge-corrected inhomogeneous estimators. The paper ends with some final conclusions.

2 Forest spatial variability: second order point process characteristics

Forest science has applied numerous methods for the statistical analysis of forest inventories (Avery and Burkhart 1983; Husch et al. 2002; West 2003). A considerable part of them belongs to spatial statistics, where the statistical analysis of point processes has played a central role (Stoyan and Penttinen 2000). Loosely speaking a spatial point process is a stochastic mechanism which generates a countable set of events x_i in a bounded region A (see, for instance, Diggle 2003). Clearly, any sequence of events, which can be seen as points scattered on a region of \mathbb{R}^d , can be explained by point process theory. Within such potential applications, the study of point occurrences in \mathbb{R}^2 , and in particular in forestry applications, has dominated point process theory (Stoyan and Penttinen 2000).

A point process Φ on \mathbb{R}^2 is characterised by its probability function $P(\Phi(A) = N)$, which is the probability of finding $N \in \mathbb{N}$ points in a given region $A \subset \mathbb{R}^2$, and its corresponding first and second order characteristics. The first order moment measure, also called the intensity measure $\Lambda(A)$, is the mean number of points contained in a given bounded region A . This measure is an important element in forest statistics, giving the mean number of trees in a forest region A . This intensity measure has a density with respect to (w.r.t.) the Lebesgue measure, the so-called intensity function $\lambda(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2$, where $\Lambda(A) = \int_A \lambda(\mathbf{x})d\mathbf{x}$ (see, for instance, Stoyan et al. 1995). Second order characteristics describe the spatial structure of point processes, and are based on the analysis of pairs of points. Although several second order characteristics have been developed based on the second order moment measures to describe point patterns (see, for instance, Stoyan et al. 1995), only the Ripley’s K function and its density w.r.t. the area have been widely used to analyse forest patterns (Stoyan and Penttinen 2000).

A point process $\Phi = \{\mathbf{x}_n\}$ is stationary if the translated process $\Phi_{\mathbf{x}} = \{\mathbf{x}_n + \mathbf{x}\}$ has the same distribution for all points $\mathbf{x} \in \mathbb{R}^2$. Whilst it is isotropic if the distribution is invariant with respect to rotation about the origin. Both features can be summarised in the idea that the spatial properties of the point process under analysis do not depend on the spatial location. For an homogeneous (stationary) point process, the intensity function reduces to a constant, the intensity λ . A standard estimator of the intensity is $\hat{\lambda} = N/|A|$, where N denotes the number of

trees contained in a forest region A , and $|A|$ is the area of this region.

A popular method to analyse spatial point patterns is based on the Ripley’s $K(r)$ function (Ripley 1976) defined as the mean number of further points within a ball of radius r centred at an arbitrary point. The K -function can be also expressed in terms of the so-called pair correlation function $g(\cdot)$ (Stoyan and Stoyan 1994)

$$K(r) = 2\pi \int_0^r g(v)v dv. \tag{1}$$

Broadly speaking, this pair correlation function indicates inhibition when $g(r) < 1$, $g(r) = 1$ denotes the Poisson case (i.e. a completely random point process) with no interaction between points, whilst $g(r) > 1$ implies point clustering. This function has been applied to a wide range of forest studies including the spatial analysis of even-aged forests (Penttinen et al. 1992; Gavrikov and Stoyan 1995), tropical forests (Pélissier 1998), tree interaction of unmanaged forests (Leemans 1991; Szwagrzyk and Czerwczak 1993; Moeur 1993) and the development of a self-thinning approach in even-aged tree populations (Gavrikov 1995). In practice, edge-corrected estimators for both the K -function and the pair correlation function are generally used (Ripley 1976; Diggle 2003).

When considering more than one species in forest stands, one can pose the question as to at what scale the tree species i segregates/aggregates from events of species j . We can answer this question by using the bivariate Ripley’s K function, defined as

$$\lambda_j K_{ij}(r) = E \left[\sum_{\mathbf{x}_i \in A \cap \Phi_i, \mathbf{x}_j \in A \cap \Phi_j} \frac{\Phi_j(b(\mathbf{x}_i, r) | \mathbf{x}_i \in \Phi_i)}{\lambda_i |A|} \right], \tag{2}$$

where $\Phi_i, \Phi_j \subset \mathbb{R}^2$, are two distinct stationary point processes with intensity functions λ_i and λ_j , respectively, and $\lambda_j K_{ij}(r)$ is, heuristically, the expected number of points of Φ_j within a ball $b(\mathbf{x}_i, r)$ centred at an arbitrary point $\mathbf{x}_i \in \Phi_i$ with radius r . Note that by symmetry $K_{ij} = K_{ji}$. The bivariate $K_{ij}(r)$ function has been widely used to analyse forest spatial patterns of (for instance): (a) young and adult trees in a southern Indian tropical forest (Pélissier 1998); (b) living and dead trees of three species in a mixed forest at Lake Duparquet (Quebec) (Park et al. 2005); and (c) young and adult trees, and between Silver fir (*Abies alba* Mill) and Norway Spruce (*Picea abies* (L.) Karst) in a case study in the Upper Susa Valley in Piedmont (Italy) (Motta and Edouard 2005).

Finally, we can also obtain a bivariate counterpart version of the pair correlation function, usually called the partial pair correlation function (see Stoyan and Stoyan 1994), also applied in numerous forest analysis (see, for

instance, Penttinen et al. 1992; Gavrikov and Stoyan 1995). Estimators for expression (2) can be found in Stoyan and Stoyan (1994).

2.1 Inhomogeneous point patterns

Real forest patterns are seldom homogeneous. Environmental heterogeneity, soil fertility and geographical configurations such as the mountain slope can promote inhomogeneous forest structures. To analyse inhomogeneous point patterns based on non-parametric summaries, Baddeley et al. (2000) formulate an inhomogeneous counterpart version of the Ripley’s K function

$$K_{inh}(r) = E \left[\sum_{\substack{\neq \\ \mathbf{x}_i, \mathbf{x}_j \in A \cap \Phi} \frac{1(\|\mathbf{x}_i - \mathbf{x}_j\| \leq r)}{\lambda(\mathbf{x}_i)\lambda(\mathbf{x}_j)|A|} \right], \tag{3}$$

where $\lambda(\mathbf{x}_i)$ is the intensity function at the spatial position $\mathbf{x}_i \in \mathbb{R}^2$ of a point process $\Phi \subset \mathbb{R}^2$ on a bounded window $A \subset \mathbb{R}^2$. An edge-corrected estimator of this function can be defined via

$$\hat{K}_{inh}(r) = \frac{1}{|A|} \sum_{\substack{\neq \\ \mathbf{x}_i, \mathbf{x}_j \in A \cap \Phi} \frac{1(\|\mathbf{x}_i - \mathbf{x}_j\| \leq r)}{\hat{\lambda}(\mathbf{x}_i)\hat{\lambda}(\mathbf{x}_j)e(\mathbf{x}_i, \|\mathbf{x}_i - \mathbf{x}_j\|)}, \tag{4}$$

where $e(\mathbf{x}_i, \|\mathbf{x}_i - \mathbf{x}_j\|)$ is the Ripley’s factor to correct for edge-effects, and $\hat{\lambda}(\mathbf{x}_i)$ is an estimator of the intensity function. A kernel-based edge-corrected estimator for the intensity function (see Silverman 1986), and posteriorly modified by Baddeley et al. (2000) to avoid biased values of \hat{K}_{inh} , is given by

$$\hat{\lambda}(\mathbf{x}) = \sum_{\mathbf{x}_i \in \Phi \cap A \setminus \{\mathbf{x}\}} \kappa_b(\mathbf{x} - \mathbf{x}_i) / C_A(\mathbf{x}_i), \tag{5}$$

where

$$C_A(\mathbf{x}_i) = \int_A \kappa_b(\mathbf{x} - \mathbf{x}_i) d\mathbf{x},$$

for any $\mathbf{x} \in \mathbb{R}^2$, and $\kappa_b(z)$ a non-negative and symmetric kernel function with bandwidth b , which has a dramatic impact on the resulting intensity surface. If b is small then the intensity surface will be concentrated around the observed events, whilst if it is large then this intensity will be close to constant. Here we consider the well-known Epanechnikov kernel, already used to analyse forest patterns (see Penttinen et al. 1992; Baddeley et al. 2000), or to be more precise the product of two such kernels,

$$\kappa_b(\mathbf{x} - \mathbf{x}_i) = \kappa_b(\|x_i - x\|)\kappa_b(\|y_i - y\|), \tag{6}$$

for $\mathbf{x} = (x,y)$ and $\mathbf{x}_i = (x_i,y_i)$, where

$$\kappa_b(|a|) = (3/4b)(1 - (|a|/b)^2)1(|a| < b),$$

and $1(\cdot)$ denotes the indicator function.

Baddeley et al. (2000) also suggest an estimator for the inhomogeneous pair correlation function

$$\hat{g}_{inh}(r) = \frac{1}{2\pi r|A|} \sum_{\substack{\neq \\ \mathbf{x}_i, \mathbf{x}_j \in \Phi \cap A} \frac{\kappa_b(r - \|\mathbf{x}_i - \mathbf{x}_j\|)}{\hat{\lambda}(\mathbf{x}_i)\hat{\lambda}(\mathbf{x}_j)e(\mathbf{x}_i, \|\mathbf{x}_i - \mathbf{x}_j\|)}, \tag{7}$$

which can be easily extended to an edge-corrected estimator for the partial (i.e. bivariate) inhomogeneous pair correlation function

$$\hat{g}_{inh}^{(ij)}(r) = \frac{1}{2\pi r|A|} \sum_{\substack{\neq \\ \mathbf{x}_i \in \Phi_i \cap A, \mathbf{x}_j \in \Phi_j \cap A} \frac{\kappa_b(r - \|\mathbf{x}_i - \mathbf{x}_j\|)}{\hat{\lambda}_i(\mathbf{x}_i)\hat{\lambda}_j(\mathbf{x}_j)e(\mathbf{x}_i, \|\mathbf{x}_i - \mathbf{x}_j\|)}, \tag{8}$$

where $\Phi_i, \Phi_j \in \mathbb{R}^2$ are two distinct inhomogeneous point processes.

3 Analysing forest patterns from replicated data

Let us consider that the forest data consists of t (replicated) point patterns, each observed on a bounded region A . Now the resulting pair correlation functions for these t plots form a sample from the theoretical overall pair correlation function, from which an overall estimate can be obtained. If the patterns are considered replicates of an underlying process, the corresponding estimates of the pair correlation function are identically distributed for a given distance r , being the average of these functions a reasonable overall estimator. Diggle (2003) considers this simple approximation to obtain overall estimates for the Ripley’s K function. Thus for the random sample of $\hat{K}(\cdot)$ functions $\{\hat{K}^{(1)}(r), \dots, \hat{K}^{(t)}(r)\}$, an overall (or pooled) estimator for these t samples can be defined as (Diggle 2003)

$$\hat{K}^P(r) = \sum_{i=1}^t N_i \hat{K}^{(i)}(r) / \sum_{i=1}^t N_i.$$

This estimator is based on the heuristic idea that $\lambda K(r)$ is the expected number of further points within a ball $b(0, r)$ centred at an arbitrary point 0 with radius r . Thus $\lambda K(r) = E[r]$, i.e. $K(r) = E[r]/\lambda$, from which this K function can be estimated by $\hat{\lambda} = \sum N_i / (r|A|)$, and $\hat{E}[r] = \sum N_i \hat{K}^{(i)}(r) / (r|A|)$ separately. The resulting estimator for the overall K is a weighted average K function which does not depend on the underlying intensity of points. Unfortunately, this weighted average estimator for the overall K is not so easily obtained for overall estimators for the inhomogeneous K function. In fact, the inhomogeneous K function not only depends on the number of points contained in each plot, but also on the spatial structure of the underlying intensity function. Assuming a common intensity surface for each replicated pattern, a very rough approximation of this estimator can be obtained by simple

averaging over the random sample of inhomogeneous K functions $\{\hat{K}_{inh}^{(1)}(r), \dots, \hat{K}_{inh}^{(t)}(r)\}$. Thus this estimator can be defined as

$$\hat{K}_{inh}^p(r) = \sum_{i=1}^t \hat{K}_{inh}^{(i)}(r)/t.$$

The same problem applies to the inhomogeneous pair correlation function (7) and inhomogeneous partial (i.e. bivariate) pair correlation function (8), for which rough overall estimators can be defined via

$$\hat{g}_{inh}^p(r) = \sum_{i=1}^t \hat{g}_{inh}^{(i)}(r)/t, \tag{9}$$

and

$$\hat{g}_{inh}^{(ij;p)}(r) = \sum_{u=1}^t \hat{g}_{inh}^{(ij;u)}(r)/t, \tag{10}$$

respectively. Once one of these rough overall estimators are obtained, we need to define their precision, which is a property of their sampling distribution. Diggle (2003) suggests the use of a bootstrap method (Efron and Tibshirani 1993) to obtain the sample variance of the overall estimate of K . Here we merely define the (empirical) standard error of this estimator based on these t plots to provide a measure of variability.

4 Testing for spatial forest inhomogeneity

The use of non-stationary point process techniques has to be motivated by the presence of non-homogeneous point patterns. Calduch (2004) computes the degree of non-stationarity of point patterns by obtaining the sum of the intersection areas of the estimated Ripley’s K function and the inhomogeneous counterpart version of this function (4), applied to the same point pattern via

$$\int_0^r \|\hat{K}(v) - \hat{K}_{inh}(v)\|dv,$$

where v is the range of possible distances. Thus for stationary point patterns this integral should be close to zero, whilst values far from zero denote non-stationarity in the point pattern. If we compare these two functions, we notice that for a given point pattern contained in a bounded region A , the only difference between them is the nature of the intensity function. Under stationarity, this intensity function is constant and equals λ (the intensity of the point process), whilst under the inhomogeneous version the point intensity depends on the spatial location, i.e. $\lambda(\mathbf{x})$. This suggests that a tentative measure of inhomogeneity for point patterns can be simply computed by comparing the

intensity function $\lambda(\mathbf{x})$ and the point intensity λ (i.e. assuming stationarity in the point pattern)

$$S = \int_A \|\lambda - \lambda(\mathbf{x})\|d\mathbf{x}. \tag{11}$$

Thus if $S = 0$, then the point pattern is stationary, whilst if $S > 0$, the point pattern can be considered non-stationary. Note that in order to compute S for a given point process Φ contained in a bounded region A with intensity function $\lambda(\mathbf{x})$, we first need to obtain the point intensity of this point process under the hypothesis of stationarity. This point intensity is that satisfying $\int_A \lambda d\mathbf{x} = \int_A \lambda(\mathbf{x})d\mathbf{x}$. To clarify this, just take into account this simple example. It is trivial to see that, for instance, the point process Φ_1 located on the unit square with intensity function $\lambda(x,y) = 70 - 20x - 20y$ is more “inhomogeneous” than the point process Φ_2 with intensity function $\lambda(x,y) = 60 - 10x - 10y$; however, under both cases we compute S assuming $\lambda = 50$, since $\int_0^1 \int_0^1 (70 - 20x - 20y)dx dy = \int_0^1 \int_0^1 (60 - 10x - 10y)dx dy = 50$.

In practice, given a point pattern, an edge-corrected estimator for the measure S can be used instead of (11)

$$\hat{S} = \int_A \|\hat{\lambda} - \hat{\lambda}(\mathbf{x})\|d\mathbf{x}, \tag{12}$$

where $\hat{\lambda} = N/|A|$ and $\hat{\lambda}(\mathbf{x})$ is given in (5). Note that when dealing with \hat{S} , i.e with $\hat{\lambda} = N/|A|$ and $\hat{\lambda}(\mathbf{x})$, we are incorporating into this measure the intrinsic stochastic variability of these estimators. Hence, two considerations are important to note here: (a) the behaviour of the measure S will depend on the correct choice and use of the intensity estimators (for example, the choice of the kernel function or the bandwidth parameter), and (b) if the theoretical measure S should be zero under stationarity, the empirical \hat{S} is not expected to be exactly zero, but should take values close to zero. We thus need to analyse the behaviour of this measure.

4.1 Statistical properties of the test

We propose to use the measure in (12) as a statistical test for inhomogeneity in a point pattern. Here the null hypothesis consists of the point pattern being stationary (homogeneous) independently of the particular spatial structure the pattern presents. Following the previous comments, under the null hypothesis, this statistical test should take values close to zero, whereas these values should be much larger under the alternative hypothesis. The statistical properties, type I error rate and power, of this test were analysed by simulation. We considered three distinct scenarios representing the three general possibilities of interaction structures in point patterns, i.e. random

(Poisson), regular (hard-core) and cluster patterns. In particular, we simulated on the circle of radius 0.5: (a) Poisson structures with intensity $\lambda = 150$; (b) Hard-core structures with threshold distance $r = 0.05$, and intensity $\lambda = 150$; and (c) Cluster structures with 37 parent events and four expected offspring per parent, giving an expected intensity of $\lambda = 150$.

For the evaluation of the type I error rate, we simulated 1,000 stationary patterns based on each of the three spatial structures considered. Power was analysed by running the test over 1,000 inhomogeneous patterns generated under the following intensity function defining the trend in the pattern $\lambda(x,y) = \exp(-2.5x - y)$. This function generates an exponential non-stationary configuration, which is encountered commonly in practice.

For each combination of spatial structure and stationarity versus inhomogeneity we obtained 1,000 values of the statistical test S . The first 500 values were used to determine the empirical distribution of S in form of histograms (see Fig. 1). The other 500 values were used to accept or reject the null hypothesis by comparing with the empirical distribution of S . By looking at Fig. 1 we can see the clear difference between the empirical distribution of S under stationarity and non-stationarity for all three spatial

structures. Table 1 reports the mean values and standard deviation of S under the considered scenarios. Note that under stationarity, the test values are clearly closer to zero with smaller standard deviation. The results concerning the type I error rate and power of the test are given in Table 2. This test presented the following type I error rates and powers: (a) $\alpha = 0.034$ (17 rejected over 500 cases) and a power of $1 - \beta = 0.992$ (496 rejected over 500 cases) for the Poisson case; (b) $\alpha = 0.032$ (16 rejected over 500 cases) and a power of $1 - \beta = 0.984$ (492 rejected over 500 cases) for the hard-core case; (c) $\alpha = 0.038$ (19 rejected over 500 cases) and a power of $1 - \beta = 0.984$ (492 rejected over 500 cases) for the cluster case.

These results suggest that our statistical test can be used as a reasonable measure of the degree of spatial inhomogeneity, and thus can be used in the practice of the statistical analysis of spatial point patterns.

5 A case study

Having presented the main statistical tools to describe the spatial structure of inhomogeneous forest patterns, we now illustrate the application of these approaches through the

Fig. 1 Frequency histograms of the empirical distribution of S under stationarity (white colour) and inhomogeneity (red colour) for the following spatial settings on the circle of radius 0.5: **a** Poisson structures with intensity $\lambda = 150$, **b** hard-core structures with threshold distance $r = 0.05$ and **c** cluster structures with 37 parent events and four expected offspring per parent

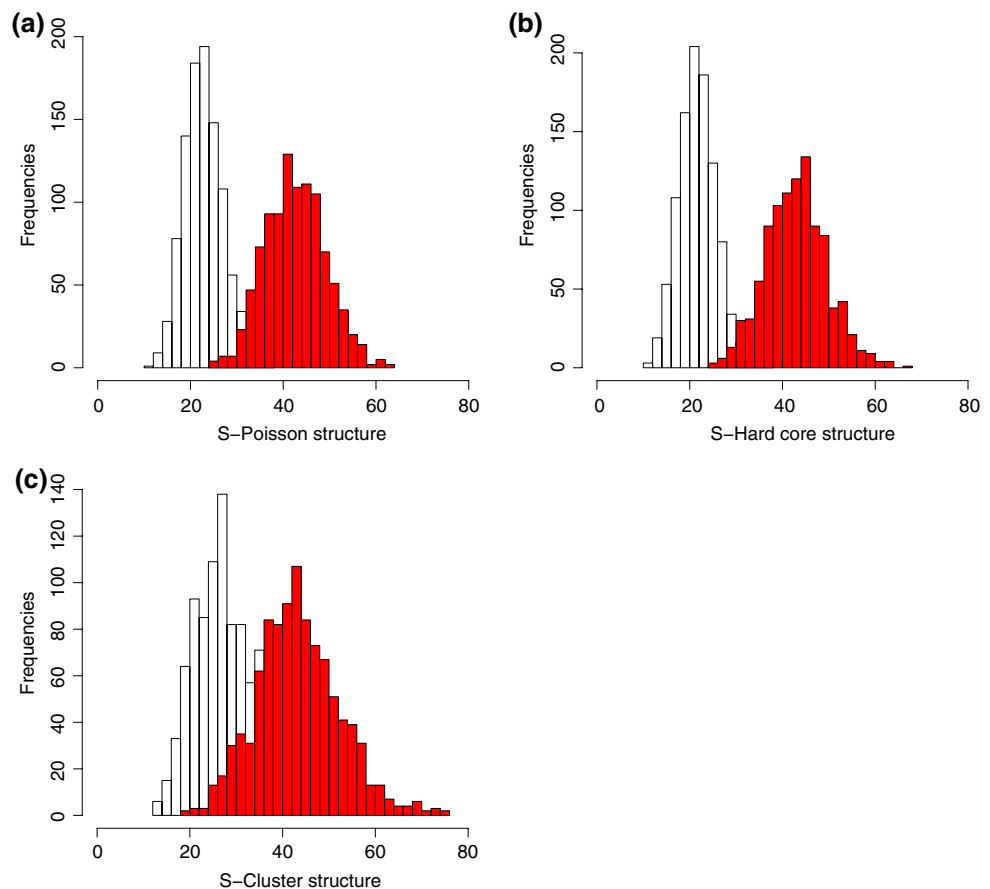


Table 1 Mean and standard deviation of the statistical test S for several spatial structures

Pattern	Structure	Mean	SD
Poisson	Homogeneous	22.82	4.13
	Inhomogeneous	42.61	6.50
Hard-core	Homogeneous	21.60	3.94
	Inhomogeneous	42.98	6.62
Cluster	Homogeneous	28.72	8.19
	Inhomogeneous	43.38	9.09

analysis of a case study in the Central Catalonia, North-East of Spain. In Central Catalonia, 45 temporary circular forest plots were established in 2006 by the Forest Technology Centre of Catalonia (CTFC) to obtain detailed information on the growing stock characteristics and the growth dynamics of *Pinus sylvestris*, *P. nigra* and *P. halepensis* pure and mixed forest stands. A total of ten, eight and nine pure stands of *P. sylvestris*, *P. nigra* and *P. halepensis*, respectively, together with 12 and 6 plots of mixed stands involving *P. sylvestris* and *P. nigra*, and *P. halepensis* and *P. nigra*, were recorded. Plots were established to contain at least 100 trees with diameter at the breast height (dbh) larger than 7.5 cm, resulting in plot sizes in the range [0.04,0.16] ha. Finally, data on tree locations, dbh, height, tree age and tree species for each plot were recorded.

Let us now characterise the degree of inhomogeneity of these forest plots. Nineteen realisations of a homogeneous Poisson process involving the same number of trees as the corresponding original patterns were simulated to obtain \hat{S} as in (12) for each realisation. Then the maximum value for each set of realisations was compared with the resulting empirical \hat{S} computed for the (original) forest pattern. If the resulting empirical \hat{S} value is larger than the maximum \hat{S} parameter obtained under simulation, the hypothesis of homogeneity (assuming a Poisson process) can be rejected at the 2.5% significance level. Although this test is defined under the hypothesis of an underlying stationary Poisson process, it could also be used under any kind of stationary point process, in which a spatial structure is present, as previously discussed. Indeed, though for any stationary

Table 2 Type I error rate and power of the test

Pattern	Structure	Accepted	Rejected
Poisson	Homogeneous	483	17
	Inhomogeneous	4	496
Hard-core	Homogeneous	484	16
	Inhomogeneous	8	492
Cluster	Homogeneous	481	19
	Inhomogeneous	8	492

point process $S = 0$, estimations of \hat{S} may depend on the stationary point process under analysis (see detailed comments in Sect. 4).

Our data analysis showed that only 40, 25 and 33% of plots of *P. sylvestris*, *P. nigra* and *P. halepensis*, respectively, could be considered non-stationary, whilst 75 and 83% of mixed stands involving *P. sylvestris* and *P. nigra*, and *P. halepensis* and *P. nigra*, respectively, should be assumed as non-stationary. There are several possible reasons why mixed stands are more inhomogeneous than pure ones. One is site variation, where each species exploits the more suitable area for survival, generating heterogeneous tree patterns. Another reason is that the regeneration history connected with distinct growth rhythms and competition effects (i.e. between and within species) promotes more inhomogeneous spatial tree configurations than under pure stands, where the nature of such competition is always between the same species. Clearly, further analysis have to be carried out to fully understand these complex forest structures.

Let us initially obtain the (global) inhomogeneous pair correlation function (9) averaged over the 45 sample plots involving all tree species. Thus, in this first analysis, we do not take into account the individual species structure. We do so to obtain the spatial structure for conifer trees in Central Catalonia. Figure 2 shows this average inhomogeneous pair correlation function together with an approximate confidence interval of this average ($\hat{g}_{inh}^p(r) \pm 2 \times \text{standard error}$) involving the 45 plots of pure and mixed stands. Note that we use the sample standard error obtained from the pair correlation functions associated to each of these 45 plots. This highlights that in average the main conifer tree species in Central Catalonia tend to be located at random. However, the study of the resulting inhomogeneous pair correlation functions for each species and the mixed stands tells a quite different story.

Figure 3 presents representative circular patterns of 20 m of radius for *P. sylvestris*, *P. nigra* and *P. halepensis* involving approximately 100 trees for each plot. We also show the resulting intensity function (grey scale image) (5), the corresponding inhomogeneous pair correlation function (7) and its respective upper and lower envelopes based on 19 complete (Poisson) randomisations. The Epanechnikov kernel with bandwidth parameter $b = 0.3$ was used to compute the intensity function in (5). This value was taken as the resulting intensity surface was (visually) fitting the observed events correctly. Whilst the bandwidth used to obtain the corresponding inhomogeneous pair correlation function was $b = 0.2/\sqrt{\lambda} \simeq 0.2/\sqrt{100/0.5^2\pi} \simeq 0.016$, as suggested by Penttinen et al. (1992) and assuming the original data rescaled to the circle of radius 0.5. Visual inspection of these three forest patterns highlights that the forest structure is clearly inhomogeneous since the

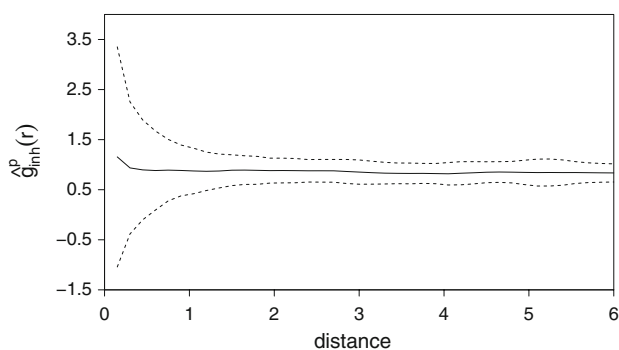


Fig. 2 Average inhomogeneous pair correlation function (9) based on 45 circular plots involving several conifer tree species and plot sizes together with its corresponding confidence interval, i.e. $\hat{g}_{inh}^p(r) \pm 2$ standard error (dashed lines)

intensity of trees depends on the spatial location under analysis. The study of the resulting inhomogeneous pair correlation functions suggests that for *P. sylvestris* the spatial structure is aggregated for short inter-tree distances ($r < 1$ m), whilst under *P. nigra* and *P. halepensis* the tree spatial configuration is regular for short inter-event ($r < 1$ m).

We now analyse the behaviour of the inhomogeneous pair correlation function averaged over all plots of the same species. Inspection of the resulting average inhomogeneous pair correlation functions (9) together with an approximate confidence interval of this average ($\hat{g}_{inh}^p(r) \pm 2 \times$ standard error) calculated over the ten, eight and nine plots for *P. sylvestris*, *P. nigra* and *P. halepensis*, respectively (see Fig. 4), confirms that whilst *P. sylvestris* tends to be aggregated for short inter-tree distances ($r < 1$ m), *P. nigra* and *P. halepensis* keep a minimum inter-event distance

($r < 1$ m) between trees. In Central Catalonia, *P. sylvestris* stands are, traditionally, regenerated by the shelterwood method (see, for instance, Smith et al. 1997). This silviculture practice may favour the creation of areas with high densities of seedlings; this may ultimately result in small clusters of trees, partially explaining the aggregated configuration for short inter-tree distances. In addition, *P. sylvestris* grows at higher elevations than *P. nigra* and *P. halepensis* where the within-plot site variation may be higher causing aggregated tree distributions. Moreover, *P. nigra* is more shade tolerant than *P. sylvestris* and may regenerate more uniformly than *P. sylvestris* under a tree canopy. The silviculture practice may also explain the regular structure for short distances in *P. nigra*. Finally, *P. halepensis* is a shade intolerant species well adapted to recurrent fires. Thus the necessity of high light intensities (i.e. direct sunlight) and the consequent self-thinning can eliminate less competitive individuals, which could explain why this species keeps a minimum inter-tree distance of around 1 m.

Finally, the study of the resulting average inhomogeneous bivariate pair correlation function (10) computed for 12 and 6 plots of mixed stands involving *P. sylvestris* and *P. nigra*, and *P. halepensis* and *P. nigra*, respectively, is shown in Fig. 5. This analysis suggests that trees of distinct species tend to be segregated from each other; this result is specially true for mixed stands of *P. halepensis* and *P. nigra*. These regular structures between trees of distinct species may be explained by within-plot site variation and inter-competition effects together with different growth rhythms of species. Another possible reason for segregation is that different species have seeds in different years, which means that locations favourable for regeneration in a

Fig. 3 Circular forest patterns of 20 m of radius with the resulting intensity functions (grey scale image; (5)) for **a** *P. sylvestris*, **b** *P. nigra* and **c** *P. halepensis*, together with the corresponding inhomogeneous pair correlation function (7) (black line) and its respective upper and lower envelopes based on 19 complete randomisations (dashed lines) for **d** *P. sylvestris*, **e** *P. nigra* and **f** *P. halepensis*

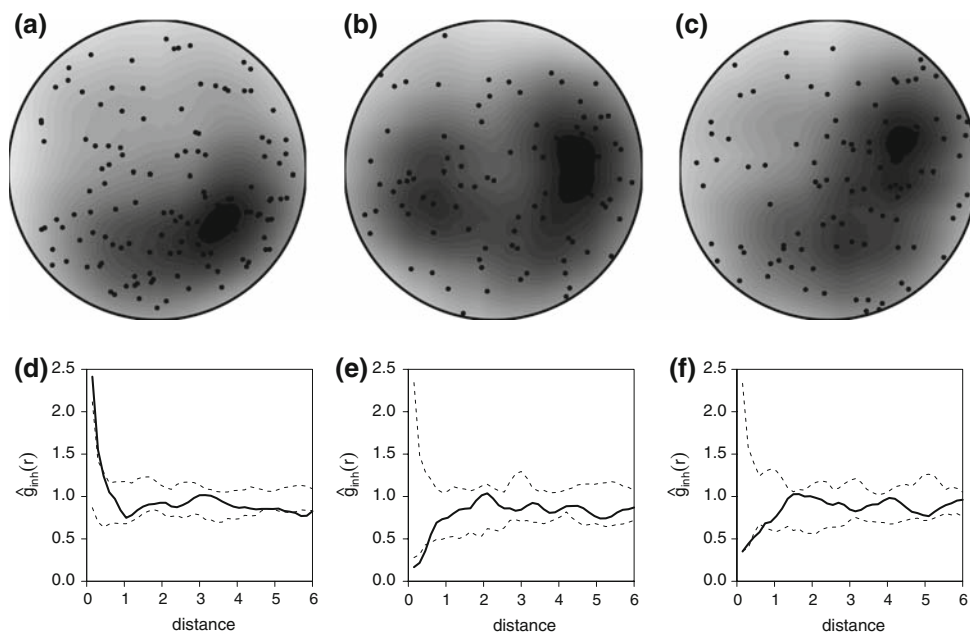
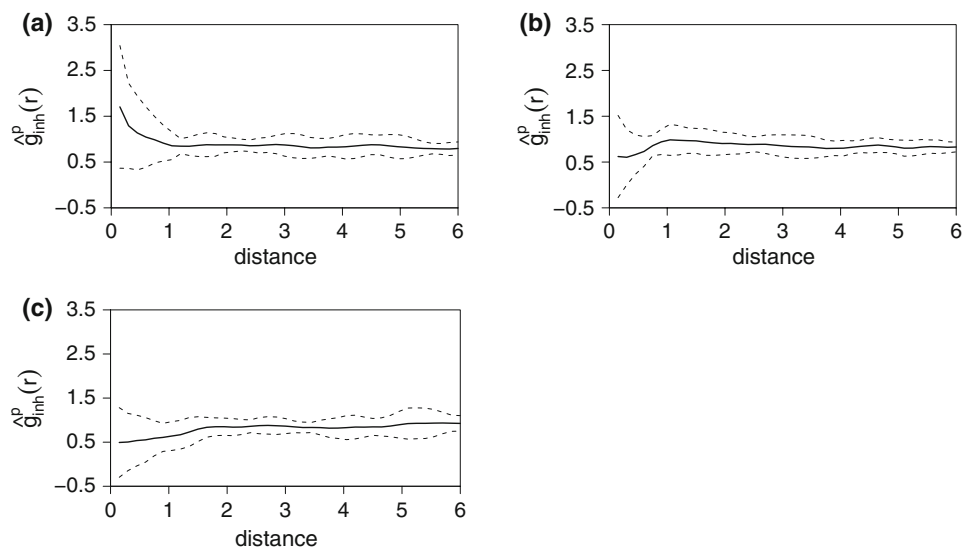


Fig. 4 Average inhomogeneous pair correlation function (9) together with its corresponding confidence interval, i.e. $\hat{g}_{inh}^p(r) \pm 2$ standard error (dashed lines) for: **a** ten plots of *P. sylvestris*; **b** eight plots of *P. nigra*; and **c** nine plots of *P. halepensis*



particular year are often regenerated and occupied by one species only.

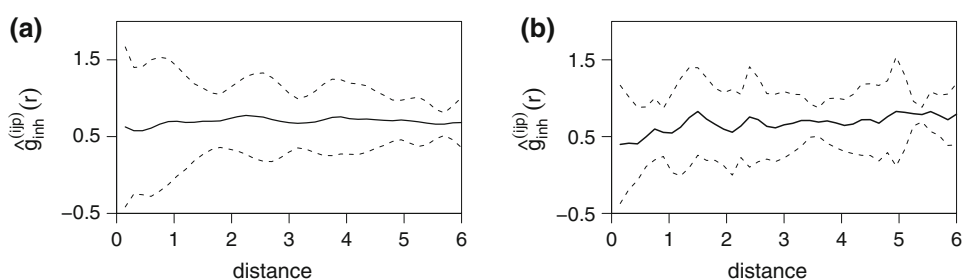
6 Conclusions and discussion

Point process theory plays a fundamental role in analysing and modelling spatial forest patterns (Stoyan and Penttinen 2000). For instance, the Ripley’s *K* function and its density w.r.t. the area, i.e. the pair correlation function, have been extensively used to analyse and characterise stationary forest configurations (see among others, Penttinen et al. 1992; Péliissier 1998; Youngblood et al. 2004). In fact, the statistical analysis of forest patterns, and in particular the description of the spatial structure, is the first step before more complicated analysis are considered, for instance model fitting. However, the use of second order characteristics (i.e. the Ripley’s *K* and the pair correlation functions) are restricted to stationary forest patterns, which regarding real forest situations are not always the case. Although the inhomogeneous counterpart version of these variability measures have been recently presented and illustrated with several practical examples by Baddeley et al. (2000), few forest studies have been performed using these inhomogeneous tools. This paper considers such

inhomogeneity characteristics to analyse the spatial structure of pure and mixed stands of conifer species in a case study in the North-West of Spain.

To justify the use of non-stationary point process statistics, we have considered a simple statistic to measure the degree of inhomogeneity based on the difference between the edge-corrected estimator of the intensity function and the estimator of the intensity (i.e. assuming a stationary point pattern). On applying this measure, we have found that whilst pure conifer plots are mainly stationary, mixed plots are clearly non-stationary. This result can be explained by environmental heterogeneity and competition effects. Further attention has to be paid on this new statistic to fully understand its behaviour. Regarding the spatial structure of these conifer species, our results suggest that whilst *P. sylvestris* tend to be aggregated for short inter-tree distances, *P. nigra* and *P. halepensis* keep a minimum inter-event distance between trees. These forest structures are most probably mainly due to site properties, competition effects, shade tolerance and silviculture practices applied in Central Catalonia for these species. Moreover, regarding the mixed stands, we found that trees of distinct species tend to be segregated from each other, indicating that small-scale site variations and inter-competition effects may affect the resulting tree spatial structure. Here

Fig. 5 Average inhomogeneous bivariate pair correlation function (10) together with its corresponding confidence interval, i.e. $\hat{g}_{inh}^{(ij;p)}(r) \pm 2$ standard error (dashed lines) for mixed stands of: **a** 12 plots of *P. sylvestris* and *P. nigra*; and **b** six plots of *P. halepensis* and *P. nigra*



the use of the average of the inhomogeneous pair correlation functions to analyse forest patterns from replicated data has provided insight into the overall forest spatial structure. However, it is clearly necessary to develop new and better strategies taking into account the non-stationary nature of such patterns to obtain such overall estimates. The understanding of real forest spatial structures will help to generate pure and mixed stands of pines which have a realistic and typical spatial configuration. These stands are required in simulators which use distance-dependent models to predict stand dynamics.

The use of inhomogeneous statistical tools has been useful to describe the spatial structure of forest patterns under analysis. However, further analysis have to be carried out considering explicit biological and forest-ecological processes to explain the nature of such complex structures. A natural further step would be to analyse not only the spatial location of trees, but also to consider characteristics associated to these spatial positions such as tree height, diameter or age in order to enable better understanding of such forest spatial structures. This could be done by generalising the pair correlation function to the marked case, and then developing the statistical test S based on the intensity of marked points.

Acknowledgments We are grateful to the Editor, AE and two anonymous referees whose comments and suggestions have clearly improved an earlier version of the manuscript. Work partially funded by grant MTM2007-62923 from the Spanish Ministry of Science and Education.

References

- Aldrich PR, Parker GR, Ward JS, Michler CH (2003) Spatial dispersion of trees in an old-growth temperate hardwood forest over 60 years of succession. *For Ecol Manage* 180:475–491
- Avery TE, Burkhart HE (1983) *Forest measurements*. McGraw-Hill, New York
- Baddeley AJ, Møller J, Waagepetersen R (2000) Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Stadistica Neerl* 54:329–350
- Calduch MA (2004) Pseudoverosimilitud e inhomogeneidad en procesos puntuales espaciales. Universitat Jaume I, Ph.D. thesis, Castellon, Spain
- Camarero JJ, Gutiérrez E, Fortin MJ (2000) Spatial pattern of subalpine grassland ecotones in the Spanish Central Pyrenees. *For Ecol Manage* 134:1–16
- Chen J, Bradshaw GA (1999) Forest structure in space: a case study of an old growth spruce-fir forest in Changbaishan Natural Reserve, PR China. *For Ecol Manage* 120:219–233
- Diggle PJ (2003) *Statistical analysis of spatial point patterns*. Hodder Arnold, London
- Efron B, Tibshirani RJ (1993) *An introduction to the bootstrap*. Chapman and Hall, London
- Ek AR, Monserud RA (1974) FOREST: a computer model for simulating the growth and reproduction of mixed-species forest stands. University of Wisconsin, Research Papers R2635
- Fortin MJ, Dale M (2005) *Spatial analysis: a guide for ecologists*. Cambridge University Press, London
- Gavrikov VL (1995) A model of collisions of growing individuals: a further development. *Ecol Modell* 79:59–66
- Gavrikov VL, Stoyan D (1995) The use of marked point processes in ecological and environmental forest studies. *Environ Ecol Stat* 2:331–344
- Husch B, Beers TW, Kershaw JA (2002) *Forest mensuration*. Wiley, New York
- Leps J, Kindlman P (1987) Models of the development of spatial patterns of an even-aged plant population over time. *Ecol Model* 39:45–57
- Leemans R (1991) Canopy gaps and establishment patterns of spruce (*Picea abies* (L.) Karst.) in two old-growth coniferous forest in Central Sweden. *Vegetation* 93:157–165
- Li L, Wang J, Cao Z, Zhong E (2007) An information-fusion method to identify pattern of spatial heterogeneity for improving the accuracy of estimation. *Stoch Environ Res Risk Assess*. doi: [10.1007/s00477-007-0179-1](https://doi.org/10.1007/s00477-007-0179-1)
- Miina J, Kolstrom T, Pukkala T (1991) An application of a spatial growth model of scots pine on drained peatland. *For Ecol Manage* 41:265–277
- Moeur M (1993) Characterizing spatial patterns of trees using stem-mapped data. *For Sci* 39:756–775
- Møller J, Syversveen AR, Waagepetersen RP (1998) Log Gaussian Cox processes. *Scand J Stat* 25:451–482
- Monserud RA, Sterba H (1996) A basal area increment model for individual trees growing in even- and uneven-aged forest stands in Austria. *For Ecol Manage* 80:57–80
- Motta R, Edouard JL (2005) Stand structure and dynamics in a mixed and multilayered forest in the Upper Susa Valley, Piedmont, Italy. *Can J For Res* 35:21–36
- Moustakas A, Hristopoulos DT (2007) Estimating tree abundance from remotely sensed imagery in semi-arid and arid environments: bringing small trees to the light. *Stoch Environ Res Risk Assess*. doi: [10.1007/s00477-007-0199-x](https://doi.org/10.1007/s00477-007-0199-x)
- Nanos N, Tadesse W, Montero G, Gil L, Alia R (2001) Spatial stochastic modelling of resin yield from pine stands. *Can J For Res* 31:1140–1147
- Onof C, Chandler RE, Kakou A, Northrop P, Wheeler HS, Isham V (2000) Rainfall modelling using Poisson-cluster processes: a review of developments. *Stoch Environ Res Risk Assess* 14:384–411
- Pacala SW, Canham CD, Silander JA (1993) Forest models defined by field measurements: I. The design of a northeastern forest simulator. *Can J For Res* 23:1980–1988
- Palahí M, Pukkala T, Miina J, Montero G (2003) Individual-tree growth and mortality models for Scots pine (*Pinus sylvestris* L.) in north-east Spain. *Ann For Sci* 60:1–10
- Park A, Kneeshaw D, Bergeron Y, Leduc A (2005) Spatial relationships and tree species associations across a 236-year boreal mixed wood chronosequence. *Can J For Res* 35:750–761
- Pélissier R (1998) Tree spatial patterns in three contrasting plots of a southern Indian tropical moist evergreen forest. *J Trop Ecol* 14:1–16
- Peng C (2000) Growth and yield models for uneven-aged stands: past, present and future. *For Ecol Manage* 132:259–279
- Penttinen A, Stoyan D, Henttonen HM (1992) Marked point processes in forest statistics. *For Sci* 38:806–824
- Pukkala T, Miina J, Kurttila M, Kolström T (1998) A spatial yield model for optimizing the thinning regime of mixed stands of *Pinus sylvestris* and *Picea abies*. *Scand J For Res* 13:31–42
- Renshaw E, Comas C, Mateu J (2008) Analysis of forest thinning strategies through the development of spacetime growth-interaction simulation models. *Stoch Environ Res Risk Assess*. doi: [10.1007/s00477-008-0214-x](https://doi.org/10.1007/s00477-008-0214-x)

- Ripley BD (1976) The second-order analysis of stationary point processes. *J Appl Probab* 13:255–266
- Ripley BD (1977) Modelling spatial patterns (with discussion). *J R Stat Soc B* 39:172–212
- Shaw D, Chen J, Freeman EA, Braun DM (2005) Spatial and population characteristics of dwarf mistletoe infected trees in an old-growth Douglas-fir-western hemlock forest. *Can J For Res* 35:990–1001
- Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London
- Smith DM, Larson BC, Kelty MJ, Ashton PMS (1997) The practice of silviculture; applied forest ecology. Wiley, New York
- Stage AR (1973) Prognosis model for stand development. USDA For Serv Res Pap INT-137
- Sterba H, Monserud RA (1997) Applicability of the forest stand growth simulator PROGNAUS for the Austrian part of the Bohemian massif. *Ecol Model* 98:23–34
- Stoyan D, Penttinen A (2000) Recent applications of point process methods in forestry statistics. *Stat Sci* 15:61–78
- Stoyan D, Stoyan H (1994) Fractals, random shapes and point fields: methods of geometrical statistics. Wiley, Chichester
- Stoyan D, Stoyan H (1998) Non-homogeneous Gibbs process models for forestry—a case study. *Biom J* 5:521–531
- Stoyan D, Kendall WS, Mecke J (1995) Stochastic geometry and its applications. Wiley, New York
- Szwagrzyk J, Czerwczak M (1993) Spatial patterns of trees in natural forests of East-Central Europe. *J Veg Sci* 4:469–476
- Teck R, Moeur M, Eav B (1996) Forecasting ecosystems with the forest vegetation simulator. *J For* 94:7–10
- West PW (2003) Tree and forest measurement. Springer, Berlin
- Wykoff WR, Crookston NL, Stage AR (1982) User's guide to the stand Prognosis model. USDA For Serv Gen Tech Rep INT-133
- Youngblood A, Max T, Coe K (2004) Stand structure in eastside old-growth ponderosa pine forests of Oregon and northern California. *For Ecol Manage* 199:191–217