

CONTRIBUTIONS OF SPATIAL POINT PROCESS MODELLING TO BIODIVERSITY THEORY

Janine ILLIAN, David BURSLEM¹

ABSTRACT

Recent decades have seen an unprecedented decline in biodiversity that has led to a growing concern about the consequences of biodiversity loss for the functioning of ecosystems. Key research in plant community ecology seeks to reveal the mechanisms that allow a large number of species to coexist and sustain biodiversity. Processes in plant communities are predominantly local and interactions take place in a spatial context. They thus need to be modelled from the individual plants' perspective.

Several ecological theories of plant species coexistence have been proposed with niche theory and neutral theory being the most prominent. They differ mainly in the extent to which functional differences between species are considered necessary for preventing competitive exclusion. This results in different predictions about interactions among the plants and between the plants and the environment.

Extensive spatially explicit data sets of plant communities have become available. This paper outlines how the theories' predictions may be assessed using spatial point process modelling and how this approach may be suitably applied to these data sets to contribute to the discussion.

Keywords : Spatial point processes, multivariate spatial point patterns, biodiversity, plant communities, tropical rainforest

RÉSUMÉ

Les dernières décennies ont connu une chute de la biodiversité sans précédent, qui soulève des inquiétudes quant à ses conséquences pour le fonctionnement des écosystèmes. Les recherches en écologie des communautés cherchent à établir les mécanismes permettant la coexistence d'un grand nombre d'espèces et le maintien de la biodiversité. Les processus mis en jeu au sein des communautés végétales sont principalement des interactions locales et prennent place dans un contexte spatial. Ils doivent donc être modélisés à l'échelle des individus. Plusieurs théories ont été proposées pour la coexistence des plantes, parmi lesquelles la théorie de la niche écologique et la théorie neutraliste sont prédominantes. Ces théories diffèrent principalement par le degré auquel les différences fonctionnelles entre espèces sont jugées nécessaires pour limiter l'exclusion compétitive. Il en résulte des prédictions différentes concernant les interactions entre espèces et entre les plantes et l'environnement. De grands jeux de données spatialisées sont à présent disponibles sur des communautés végétales, comportant la localisation de chaque plante. Cet article discute comment les prédictions des différentes théories peuvent être évaluées

1. School of Biological Sciences, University of Aberdeen, UK, j.illian@abdn.ac.uk

à l'aide de modèles de processus ponctuels et comment l'approche peut être appliquée à ces jeux de données pour contribuer à la discussion.

Mots-clés : Processus ponctuels spatialisés, semis de points multivariés, biodiversité, communautés végétales, forêt tropicale humide

1. Introduction

Ultimately, the earth as a system is dependent on the functioning of natural and managed ecosystems since it is regulated by the biogeochemical processes derived from them (Loreau *et al.*, 2001). Recent decades have seen an increasing decline in species' biodiversity as a result of human interference (Cardinale *et al.*, 2004 ; Regan *et al.*, 2001). The potential ecological consequences of biodiversity loss have led to a growing concern about the future survival of ecosystems and their functioning (Mouquet *et al.*, 2002). Consequently the relationship between biodiversity and ecosystem functioning constitutes a major scientific issue today (Loreau, 2000 ; Tilman *et al.*, 1996, 2001 ; Hector *et al.*, 1999). However, understanding the impact of biodiversity loss on function requires an understanding of the processes that structure communities and the mechanisms that sustain biodiversity (Bell, 2001 ; Hubbell, 2001 ; Chave, 2004; Condit *et al.*, 2002 ; Duivenvoorden *et al.*, 2002). These mechanisms are poorly understood.

Key research in community ecology thus seeks to reveal the mechanisms that allow a large number of species to coexist (Murrell *et al.*, 2001; Loreau *et al.*, 2001). Coexistence is determined in part by the inter- and intra-specific interactions in a community (Durrett and Levin, 1998). Since individual plants interact mainly with their neighbours (Tilman, 1994; Dieckmann *et al.*, 2000; Prues and Law, 2003; Stoll and Weiner, 2000), interactions between plants in plant communities typically take place in a spatial context and hence current modelling approaches consider individuals in spatially explicit models (Chesson, 2000; DeAngelis and Gross, 1992; Huston *et al.*, 1988; Judson, 1994). It has long been recognised that processes in plant communities are predominantly local and need to be modelled from the "plant's eye view" (Turkington and Harper, 1979; Hubbell, 1979; Law *et al.*, 2003), i.e. from the point of view of every individual in a plant community, thus taking local growing conditions as well as local competition into account. Spatial point processes are stochastic models that describe the spatial pattern formed by the locations of objects in two- or more dimensional space. The locations are regarded as points and the aim is to describe the pattern formed by these points. In this paper we aim to outline why we believe that spatial point process models may serve as an appropriate tool for the analysis of pattern formations in ecological communities and illustrate how they may contribute to debates in biodiversity theory by enabling tests of competing theories of plant species coexistence.

We believe that spatial point process methods are an appropriate statistical tool for analysing the relationship between underlying ecological processes and the resulting spatial pattern. Being based on individual plants, the approach takes the plant's eye-view into account and has the capacity to summarise the observed pattern with a few parameters whilst still taking the individuals as well as the underlying processes into account.

However, up to now spatial point process methodology has only rarely been applied to model entire plant communities, and has merely been used in a descriptive manner (Diggle, 2003; Mateu *et al.*, 1998). Recent developments in both spatial point process theory and ecology have made the methods more applicable in this context, in particular through the development of multivariate methods for spatial point processes and parsimonious modelling techniques (Illian *et al.*, 2004, 2005, 2006), see below.

This paper is organised as follows. Section 2 presents a short overview of the debate concerning the maintenance of species richness in plant communities, Section 3 briefly introduces spatial point process theory, Section 4 details current and future contributions of the methodology to resolve the ongoing discussion in biodiversity theory.

2. Biodiversity – the theoretical debate

Although many theories of plant species coexistence have been proposed, the principle criterion that distinguishes between them is the extent to which functional differences between species (their individual niches) are necessary and important for preventing competitive exclusion. We briefly describe two of the main theories.

2.1. Niche theory

The term ecological niche, which originates in Elton (1927) and Grinnell (1917) and more formally in Hutchinson (1957, 1959), refers to the multi-dimensional summary of a species' ecological attributes, including its abiotic tolerances, its maximum relative growth rate, its phenology, its susceptibility to enemies and its relative ability to compete with other plant species (Crawley, 1997).

Classically, plant ecological theory has maintained that different species occupy different niches, i.e. that they have adapted to survive in a specialised environment by exploiting the available resources. Speciation prevents inter-specific competition from driving competitive exclusion and thus promotes long-term coexistence of several species in the same area. Niche theory states that species-specific differences influence the population dynamics and therefore the behaviour of the whole community (Purves and Pacala, 2005). As a consequence, biodiversity and functioning are directly linked from a niche theoretical perspective; loss in biodiversity directly influences a community's functioning.

2.2. Neutral theory

Alternative models, of which Hubbell's (2001) neutral theory is the most prominent, have relaxed or abandoned this assumption of individual species' niche partitioning and model species persistence as a balance between the dynamics of extinction, immigration from the metacommunity and species accumulation through speciation. It assumes that all species and thus all individuals in a community are equivalent, such that they are interchangeable, independent of environmental conditions and of space and time (Purves and Pacala, 2005; Chave, 2004). The theory acknowledges that while there are differences in species' properties, these do not affect the population dynamics and hence have no impact on the behaviour of a community or its biodiversity (Hubbell, 2001). In fact, in a completely neutral community all but one species can be eliminated without affecting the biogeochemical functioning of the community.

The theory postulates that species abundances follow a random walk or *drift*, with equal per capita probabilities of birth and death. The only constraint here is that the total number of individuals over all species in the community is constant (de Mazancourt, 2001). There is no superior competitor and the probabilities of death and birth, dispersion patterns and speciation are equivalent for all species. The community is saturated with individuals; as soon as one individual dies its place will be taken up by another individual, leading to strong competition (Gaston and Crown, 2005). Despite the assumption of ecological equivalence, simulated communities whose dynamics are governed only by ecological drift have emergent properties that closely approximate those of real communities (Hubbell 2001), which lends credibility to the idea that species differences need not be important in community assembly.

2.3. Niche or neutrality?

The relative importance to species coexistence of niche partitioning determined by differences among species and ecological drift due to species' equivalence is still unclear. Both theories are supported by substantial bodies of evidence (Bell, 2001; Hubbell, 2001; Chave, 2004; Condit *et al.*, 2002; Duidenvoorden *et al.*, 2002). The neutral theory assumes that all species are ecologically equivalent, and therefore infers that biodiversity must be uncoupled from ecosystem functioning. Testing the relative importance of niche partitioning and neutrality is therefore very important to predicting how ecosystem services such as productivity, resilience to disturbance and carbon storage will respond to reductions in biodiversity (Purves and Pacala, 2005).

Testing the relative importance of the various mechanisms proposed for the maintenance of species richness in plant communities has been complicated by their overlapping predictions when applied to real plant communities, and by a lack of consensus over an appropriate framework for the analyses. We propose that spatial point process methodology should be considered as one candidate analytical framework for testing these competing theories.

Above all, spatial point process methodology is suitable for application to the problem of species coexistence in plant communities because the alternative mechanisms reviewed above all have embedded predictions about the spatial arrangement of individuals in communities, or of interactions between neighbours in space. For example, most definitions of a plant's niche infer non-randomness in its association with underlying biophysical factors such as microclimate, soil conditions or parent material (Grubb, 1977), and interactions that define an appropriate abiotic and biotic environment for regeneration (Schoener, 1989).

3. Spatial point processes

This section will provide a very brief introduction to spatial point processes. A very rigorous theoretical introduction to the general theory of point processes may be found in Daley and Vere-Jones (1988). Overviews of the theory of spatial point processes are given in Møller and Waagepetersen (2003a,b); van Lieshout (2000); Stoyan and Stoyan (1994); Stoyan *et al.*(1995); Stoyan and Penttinen (2000); Diggle (2003).

3.1. Definition

A spatial point process X is defined as follows.

DEFINITION 1. — For each Borel set $B \subset \mathbb{R}^2$, let $\phi_X(B)$ be the number of points of X in B . Here, we identify a point configuration with a counting measure ϕ_X on Borel sets on \mathbb{R}^2 .

Let N be the set of all such measures. On N define \mathcal{N} as the smallest σ -algebra generated by sets of the form

$$\{\phi \in N : \phi(B) = n \quad n \in \{0, 1, 2, \dots\}, B \text{ some bounded Borel set}\}.$$

Let (Ω, \mathcal{A}, P) be some probability space. A spatial point process X may then be regarded as a measurable mapping from (Ω, \mathcal{A}) into (N, \mathcal{N}) , i.e. as a random variable. A spatial point pattern x is then a realisation of this random variable.

To avoid unnecessary notation we will not distinguish between X and the measure ϕ_X defined by it, i.e. $X(B)$ denotes the number of points X has in B . In the applications considered here, the process X lives in some subset W of \mathbb{R}^2 and patterns are only observed in a bounded area $S \subset W$. Even though S can be of a very general form, we assume here that without loss of generality S is a rectangular set $S \subset W \subset \mathbb{R}^2$. Individual points in X will typically be denoted by ξ and η . Locations in S which may or may not coincide with a point in X will be denoted by u .

Note that more general spaces may be considered such as the \mathbb{R}^d or other metric spaces equipped with a metric $d(\cdot, \cdot)$ which are Polish, i.e. complete and separable. For details see Daley and Veres-Jones (1988).

Note also that we assume X to be simple, i.e. not more than one point may occur in any location. In situations where additional data exist on the objects that form the spatial point pattern under investigation, this is denoted a marked point process.

DEFINITION 2. — Let Z be a simple point process in \mathbb{R}^2 . Attach a random mark $m_\xi \in \mathcal{M}$, where \mathcal{M} is some mark space, to each point $\xi \in Z$. This yields a marked point process

$$X_m = \{(\xi, m_\xi) : \xi \in Z\}.$$

In most applications, the mark space \mathcal{M} is a subset of \mathbb{R}^d with $d \geq 1$, but more general mark spaces may be considered, see Schlather (2001) or Stoyan and Stoyan (1994).

Note that if $\mathcal{M} = \{1, \dots, k\}$, X is a multi-type point process with k different types of points. Note further that a multi-type process can also be regarded as a k -tuple of different subprocesses (X_1, \dots, X_k) .

3.2. The spatial Poisson process

A very simple point process model is the spatial Poisson process.

DEFINITION 3. — Let ρ be a locally finite and diffuse measure defined on the Borel sets in S , i.e. $\rho(B) < \infty$ for all bounded Borel sets B on S and ρ has no mass at any point in S . Then the Poisson process on S with intensity measure ρ written as

$$X \sim \text{Poisson}(S, \rho)$$

has the following properties:

- (1) For any Borel set B in S the cardinality of B , $X(B)$, follows a Poisson distribution with mean $\rho(B)$.
- (2) For any disjoint Borel sets $B_1, \dots, B_{n_0} \subseteq S$ with an arbitrary $n_0 \geq 2$, $X(B_1), \dots, X(B_{n_0})$ are independent.

If the measure ρ is given by a density λ with respect to the Lebesgue measure then we call λ the intensity function of the process and write $X \sim \text{Poisson}(S, \lambda)$ instead of $X \sim \text{Poisson}(S, \rho)$.

Note that Poisson process

$$X \sim \text{Poisson}(S, \mu \cdot \lambda_0),$$

with constant intensity λ_0 is termed the homogeneous Poisson process, where μ is the Lebesgue measure. The homogeneous Poisson process is a suitable null model to describe complete spatial randomness (Stoyan *et al.*, 1995) and is used as a reference for the exploratory analysis of a spatial pattern.

It is often convenient to express a more general point process model in terms of its density, notably for the class of Markov processes, see below, with regard to the unit rate homogeneous Poisson process in accordance with the Radon-Nikodym theorem (Møller and Waagepetersen, 2003b).

Note, that general Poisson processes are always absolutely continuous with respect to the standard or unit rate Poisson process, i.e. the Poisson process with constant intensity $\lambda_0 = 1$, when defined on a bounded subset S of \mathbb{R}^d (Møller and Waagepetersen, 2003b).

For a general Poisson process

$$f(x) = \prod_{i=1}^{n(x)} \lambda(x_i) \exp\left(-\int_S [\lambda(\mu) - 1] d\mu\right)$$

is the density function (Møller and Waagepetersen, 2003b). We will present further examples below.

3.3. Summary characteristics

3.3.1. First order summary characteristics

For a point process X the intensity measure is given by

$$\Lambda(B) = E[\phi(B)], \quad \text{for any Borel set } B,$$

where ϕ is a counting measure as defined above and $E[\cdot]$ denotes the expected value. If Λ is absolutely continuous with regard to the Lebesgue measure, a density function, the intensity function $\lambda : B \rightarrow \mathbb{R}^+$ exists, such that

$$\Lambda(B) = \int_B \lambda(x) dx.$$

Details on the estimation of first order summary characteristics may be found e.g. in Stoyan *et al.* (1995); Møller and Waagepetersen (2003b).

3.3.2. Second order summary characteristics

Several second order summary characteristics have been proposed in the literature (see for example Ripley (1976)). Second order summary characteristics suitable for homogeneous data such as Ripley's K -function have become standard tools in applications (Diggle, 2003). We thus only discuss the inhomogeneous K - and L -function as well as the inhomogeneous pair correlation function as introduced in Baddeley *et al.* (2000) as these are less well-known. The inhomogeneous K -function is defined as

$$K_{\text{inhom}}(r) = E \sum_{\xi \in X, \eta \in X, \xi \neq \eta} \mathbf{1}[\|\xi - \eta\| \leq r] / (\lambda(\xi)\lambda(\eta)). \quad (1)$$

Often, the variance stabilising L -function is used instead, where

$$L_{\text{inhom}}(r) = \sqrt{\frac{K_{\text{inhom}}(r)}{\pi}}.$$

Another second order summary characteristics commonly used is the pair correlation function. For two Borel sets B_1 and B_2 we define the second order factorial moment measure $\mu^{(2)}$ as

$$\mu^{(2)}(B_1 \times B_2) = E \sum_{\xi \in X, \eta \in X, \xi \neq \eta} \mathbf{1}[\xi \in B_1, \eta \in B_2].$$

If $\mu^{(2)}$ has a density function $\rho^{(2)}$, i.e. if

$$\mu^{(2)}(B_1 \times B_2) = \int_{B_1} \int_{B_2} \rho^{(2)}(\xi, \eta) d\xi d\eta$$

then $\rho^{(2)}$ is called the second-order product density. The pair correlation function is defined as

$$g(\xi, \eta) = \frac{\rho^{(2)}(\xi, \eta)}{\lambda(\xi)\lambda(\eta)}.$$

Details on the estimation of second order summary characteristics may be found e.g. in Stoyan *et al.*(1995); Baddeley *et al.*(2000); Møller and Waagepetersen (2003b).

Note that the second order summary characteristics may be used to distinguish clustered, random and regular patterns. For a Poisson process we have $K(r) = \pi r^2$, $L(r) = r$ and $g(r) = 1$. Thus, if $K(r) > \pi r^2$ at close distances we have a clustered pattern, $K(r) < \pi r^2$ indicates a regular pattern and analogously for the other characteristics.

3.4. More general models

The concept of the Poisson process may be generalised to provide more flexible models that are more suitable for applications.

3.4.1. Markov point processes

The class of spatial Markov point processes models patterns exhibiting aggregation (or inhibition) due to interaction between points (van Lieshout, 2000). A special case of Markov point processes are pairwise interaction processes. Its density is of the following form:

$$f(x) = \alpha \prod_{i=1}^{n(x)} \beta(\xi) \prod_{\xi, \eta \in X, \xi \neq \eta} h(\|\xi - \eta\|), \quad (2)$$

where α is a normalising constant, $\beta(\cdot)$ is a function describing the intensity of the process and $h(\cdot)$ is a non-negative interaction function.

A number of interaction functions have been considered in the literature; van Lieshout (2000) gives a detailed overview of various choices of interaction functions. For restrictions on the choice of interaction functions see, for example Diggle (2003).

3.4.2. Neyman-Scott processes

Neyman-Scott processes, are a class of spatial point processes that describe a "mother-daughter" relationship between points (Diggle, 2003; Stoyan *et al.*, 1995). These processes are constructed in two steps.

1. "Mother" points are generated from a Poisson process with intensity function λ .
2. For each mother, a random number of "daughters" is generated, where the number of offspring is independently identically realised from a univariate distribution for each parent, and the locations of the daughters follow a bivariate distribution around the mother points.

The locations of the points in the daughter process form the Neyman-Scott process. Examples of Neyman-Scott processes include the Matérn cluster process, where the locations of the daughter points are uniformly distributed on the area of a circle with radius r_T around the mother points. Similarly, the Thomas processes is a point process model, where the locations of the daughter points follow a bivariate normal distribution with mean 0 and variance-covariance matrix $\sigma_M^2 I$ around the mother plants, where I denotes the 2×2 unit matrix (Møller and Waagepetersen, 2003b). Neyman-Scott processes are appealing in an applied context since closed form expressions of the theoretical summary statistics are known for some of these models. This facilitates parameter estimation (Diggle, 1983).

3.4.3 Cox processes

Cox processes are a class of models describing aggregation or clustering resulting from unobserved environmental variability influencing the location of points in a process. This variability is assumed to be a stochastic process in itself leading to the Cox processes being called "doubly-stochastic" processes, a term introduced in Cox (1955). They are constructed such that the resulting process is a Poisson process, given the underlying random process.

Formally, this yields the following definition:

DEFINITION 4. — *Let $Z = \{Z(u) : u \in S\}$ be a non-negative locally finite random field on S . If the conditional distribution of a spatial point process X given Z , denoted by $X|Z$ has $X|Z \sim \text{Poisson}(S, Z)$, then X is called a Cox process conditional on Z .*

Remark. — For the exact definition of a random field and its properties, see for example Adler (1981). For our purposes it suffices to say that $Z(u)$ is a random variable for all $u \in S$. Hence, we can regard Z as a random intensity function. Under certain regularity conditions, Z can be associated with a corresponding measure Λ , say, i.e. a Cox process can equivalently be defined in terms of a random measure, for details see Møller and Waagepetersen (2003b).

3.5. Multivariate methods for spatial point processes

Recently, attempts have been made to develop multivariate approaches and, most notably, principal component analysis for spatial point pattern data (Illian *et al.*, 2004, 2005). These use methods from functional data analysis on the second order summary statistics. We briefly describe the approach here.

3.5.1. Functional principal component analysis

For a more detailed introduction to functional data analysis see Ramsay and Silverman (1994, 2002). Functional data analysis operates on functional data, i.e. observations that are functions interpreted as single entities rather than as consecutive measurements. Generally, the record of a functional observation x consists of n pairs $(t_j, y_j), j = 1, \dots, n$, where y_j is an observation of the function $x(t_j)$ at time t_j . Since the functions are usually observed at a finite number of values of t_j only, interpolation or smoothing techniques have to be applied to yield a functional representation of the data. This is typically done using the basis function method where a function is represented by a linear combination of K unknown basis functions ψ_k , i.e. $x(t) = \sum_{k=1}^K c_k \psi_k(t)$. The coefficients c_k of the expansion are determined by minimising the least square criterion

$$LS(y|\mathbf{c}) = \sum_{j=1}^n [y_j - \sum_{k=1}^K c_k \psi_k(t_j)]^2.$$

Now, in the context of functional principal components analysis (FPCA) we consider function values $u_i(t)$ and define

$$f_i = \int \beta(t) u_i(t) dt,$$

where $\beta(t)$ is a weight function, and maximise $N^{-1} \sum_i^N f_{i1}^2$ under the constraint $\|w_1\|^2 = \int w_1(t)^2 dt = 1$ and have an eigenequation $\int v(t, s) w(s) ds = \lambda w(t)$ with variance-covariance function $v(t, s) = N^{-1} \sum_{i=1}^N u_i(t) u_i(s)$. The solution to this eigenequation with largest eigenvalue solves the maximisation problem.

In the subsequent p steps maximise $N^{-1} \sum_i^N f_{ik}^2$, where $k = 1, \dots, p$, subject to $\|w_i\|^2 = \int w_i(t)^2 dt = 1$. Further analysis will examine the scores f_{ik} for each of the original smoothed curves on the first p principal components, with $p \ll k$ in connection with an interpretation of the shape of these principal components.

3.5.2. Principal component analysis for spatial point patterns

To apply the above approach to spatial point pattern data consider the following. Let Z be a spatial point process on \mathbb{R}^2 . Let X be a multitype point process $X = \{(\zeta, m_\zeta) : \zeta \in Z\}$ with $m_\zeta \in \mathcal{M}$ and $\mathcal{M} = \{1, \dots, k\}$ a set, where no other marks are available, and subprocesses $X_i \subset X$ with $X_i = \{(\zeta, m_\zeta) : \zeta \in Z \text{ and } m_\zeta = i\}$ and $i = 1, \dots, k$. Consider a realisation x of X . We use second-order summary statistics, in particular pair correlation functions, to characterise the spatial behaviour of the individual subpatterns x_i .

We smooth the estimated second-order summary statistics using cubic B-splines. We subsequently perform a functional PCA on the smoothed functions. Through this, the subprocesses may be grouped on the basis of their scores on the (functional) principal components and those aspects of spatial behaviour that vary most among the species are revealed (Illian *et al.*, 2004, 2005). To illustrate this Figure 1 shows the first two principal components resulting from a functional PCA based on pair correlation functions for a data set derived from a species-rich shrub community in western Australia (see Armstrong (1991); Illian *et al.* (2005) and Section 4.1 for more details on the data set). The plot indicates that the species' spatial behaviour varies most with regard to presence and absence of clustering at close distances as described by the first PC (Illian, 2006).

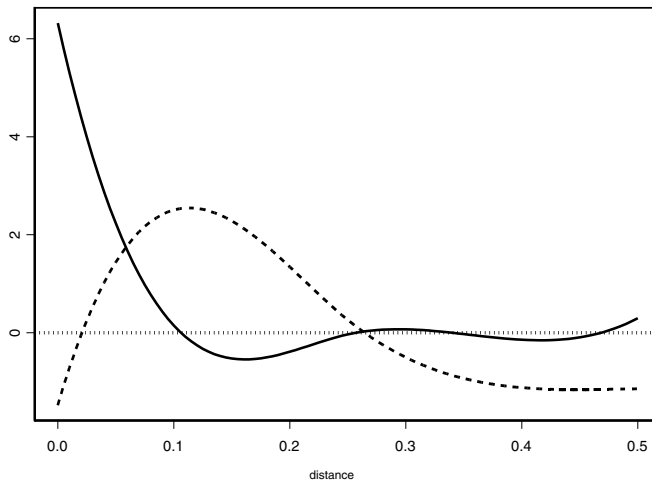


FIG 1. — Plot of the first two principal component functions for the Australian data set; the full line is the 1st PC, the dotted line is the 2nd PC.

Figure 2 plots the scores on the first principal components for each of the species. This yields a characterisation of the species' spatial behaviour in two-dimensional space.

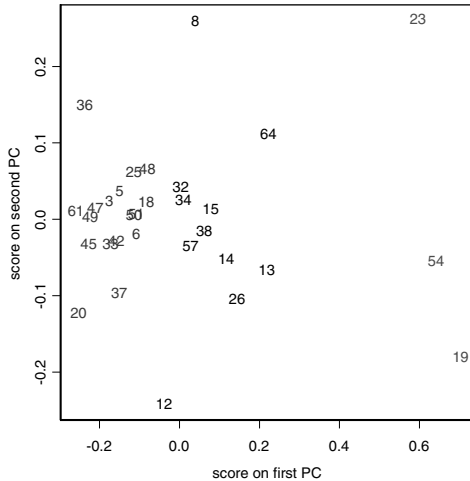


FIG 2. — Plot of the scores on the first two principal components for the Australian data set; numbers indicate species numbers as in Illian (2006).

4. Spatial point processes and biodiversity theory

Until recently, ecologists lacked substantial datasets that were suitable for testing the explicit predictions of the competing theories of biodiversity. A suitable dataset is one in which the locations of individual plants are spatially referenced, identified to species, and censused over ecologically meaningful timescales. To our knowledge, the most substantive current datasets that possess these characteristics have been obtained from the network of tropical forest plots coordinated by the Center for Tropical Forest Science of the Smithsonian Tropical Research Institute. For example, on the first plot in the network, the 50 ha forest dynamics plot on Barro Colorado Island (BCI) in Panama, Hubbell & Foster (1986a,b, 1987) first recorded the locations and sizes (diameter at breast height) of 235,349 individuals of 304 tree species in 1982 (Condit *et al.*, 2000). Since then, additional environmental data have been collected, including information on soil properties, water availability and topography. The plot has been recensused every five years since 1985 (Condit *et al.*, 2002). Similar plots have now been established to create a network of 16 tropical forest plots in 14 countries that all follow standardised protocols for data collection.

A large number of studies has been conducted using the data collected from these plots (summarised in Losos & Leigh 2005), but to date spatial point process methodology has not been applied to plant community data sets, other than in a descriptive way (Coomes *et al.*, 1999). We advocate the application of spatial point process modelling techniques to the spatially explicit tropical forest data because of their potential ability to test the spatial predictions of the competing models of species coexistence. We focus on three subject areas where application of spatial point process modelling techniques might

stimulate exciting new breakthroughs in our understanding of plant species coexistence.

4.1. Interaction strength and neighbourhood identity

We preface this discussion with the observation that the terms "interaction" and "interaction strength" have different meanings in theoretical ecology and spatial statistics. The interaction parameters in spatial point process models refer to the probability of individuals of the same or of different species occurring in close proximity. This should not be confused with the term interaction in ecology, which refers to a change in state of one or more organisms over time as a result of a process in which they participate.

Within a species the interaction parameters in a spatial point process model reflect the deviation of a particular pattern from a random pattern, i.e. it's tendency to be either aggregated or repulsive. Between species they reflect the deviation of independence of the species' patterns, i.e. the tendency of individuals from different species to appear more frequently or less frequently in close proximity than if the pattern were independent. One might term the concept that these parameters represent "spatial interaction"; we shall refer to it as interaction from now on. The strength, direction, range and significance of this interaction may be assessed (van Lieshout, 2000), i.e. the models may be applied to inform on inter-individual interaction in a given community as reflected by the spatial pattern.

Exploratory data analysis and principal component analysis for spatial point patterns of a species-rich shrub community in western Australia (Illian *et al.*, 2005) have revealed that intra-specific interaction varies among species. Illian *et al.* (2006) present a spatial point process model for the Australian data using the following approach.

The species in the data set basically have two different fire regeneration strategies, they either survive a fire underground (resprouter species) or shed their seed triggered by a fire event (seeder species). The seeders are modelled conditional on the resprouters to yield an improved understanding of the nature of the inter-species interactions among them.

Denote by W the 22 m by 22 m plot where the plants were recorded, x_1, \dots, x_{19} the observed point patterns for the 19 resprouters, y_1, \dots, y_5 the observed point patterns for the 5 seeders, and $X_1, \dots, X_{19}, Y_1, \dots, Y_5$ the corresponding spatial point processes, i.e. here each X_j or Y_i is considered to be a random finite subset of W .

Conditional on $X_1 = x_1, \dots, X_{19} = x_{19}$ assume that Y_1, \dots, Y_5 are independent Poisson processes, i.e. that there is no intra-species interaction for the seeder species. Also assume that each Y_i has intensity function

$$\lambda(\xi|x, \theta_i) = \exp(\theta_i s(\xi|x)^T), \quad \xi \in W, \quad (3)$$

where $x = (x_1, \dots, x_{19})$ is the collection of all 19 resprouter point patterns; $\theta_i = (\theta_{i0}, \dots, \theta_{i19})$ is a vector of parameters, where $\theta_{i0} \in \mathbb{R}$ is an intercept and for $j = 1, \dots, 19$, $\theta_{ij} \in \mathbb{R}$ controls the influence of the j th resprouter on the

i th seeder (a positive value of θ_{ij} means a positive/attractive association; a negative value of θ_{ij} means a negative/repulsive association); $s(\xi|x) = (1, s_1(\xi|x), \dots, s_{19}(\xi|x))$ with

$$s_j(\xi|x) = \sum_{\eta \in x_j} h_\eta(\|\xi - \eta\|), \quad j = 1, \dots, 19,$$

where $\|\cdot\|$ denotes Euclidean distance; and h_η is a smooth interaction function given by

$$h_\eta(r) = \begin{cases} (1 - (r/R_\eta))^2 & \text{if } 0 < r \leq R_\eta \\ 0 & \text{else} \end{cases}$$

for $r \geq 0$, where $R_\eta \geq 0$ defines the radii of interaction of a given resprouter at location η .

Thus, given the resprouters x the number N_i of points in Y_i is Poisson distributed with mean value $\int_W \lambda(\xi|x, \theta_i) d\xi$, and if we also condition on N_i , the points in Y_i are independent and identically distributed with a density proportional to $\lambda(\xi|x, \theta_i)$. It follows that the log likelihood function based on the 5 seeder point patterns $y = (y_1, \dots, y_5)$ is

$$l(\theta, R; y|x) = \sum_{i=1}^5 \left[\theta_i \sum_{\xi \in y_i} s(\xi|x)^T - \int_W \exp(\theta_i s(\xi|x)^T) d\xi \right], \quad (4)$$

where $\theta = (\theta_1, \dots, \theta_5)$ is the vector of all 100 parameters θ_{ij} and R is the vector of all 3168 radii R_η , $\eta \in x_j$, $j = 1, \dots, 19$ of the 3168 resprouter plants from the 19 species.

The results indicate that interaction strength and direction vary with the identity of the species concerned, since negative, positive and no interactions were found. The findings support the prediction of niche theory that the identity of neighbouring plants influences spatial structure within a population. However, both significant and non-significant interaction parameters were found when all species pairs were analysed, therefore it would be premature to conclude that niche theory is supported universally by this community.

4.2. Microhabitat specialisation

Evidence of microhabitat specialisation provides support for niche-assembly models of community dynamics, as evidence of adaptation to specific environments challenges the assumption of ecological equivalence that is inherent in neutral theory (Harms *et al.*, 2001). A spatial point process analysis is able to detect these aggregations and models the spatial pattern formed by the individuals taking local environmental conditions into account. By merely analysing the spatial point pattern of the species alone it is not possible to distinguish between aggregation due to dispersal, interactions or environmental heterogeneity unless data on the environmental conditions in specific locations are available.

The point pattern data for the Australian shrub community lack supplementary data on environmental covariates, and an analysis of microhabitat specialisation is therefore not possible. Nevertheless, for the Australian data set one may assume that the growth conditions are homogeneous throughout the plot and interpret the resulting pattern as resulting primarily from intra- and inter-specific interactions. Using principal component methodology, groups of species with similar intra-specific behaviour could be identified. More importantly, those aspects of spatial behaviour that vary most among the species could be established. This is interesting because the data set is derived from a truly ancient community, which has established its dynamics over thousands of generations in an area that has undergone substantial climatic changes (Dixon 2005). One may assume that we are dealing with a very stable community that has adapted to the given circumstances in an optimal way resulting in the species developing niche behaviour that is crucial for stable coexistence. It could thus be concluded that the spatial behaviour at small distances constitutes a niche-behaviour, which may support coexistence of the large number of species in the area under investigation.

By contrast, the spatially explicit tropical forest tree position data have linked data-sets on elevation, aspect, and in some cases soil conditions (reviewed in Losos & Leigh 2005) and therefore provide an opportunity to test for associations between species distribution and microhabitat variation. Waagepetersen (in press) determined the functional relationship between the intensity of two species, *Ocotea whitei* and *Beilschmiedia pendula* (both Lauraceae), and elevation and gradient of the slope on Barro Colorado Island, Panama. Here, an inhomogeneous Neyman-Scott process was applied in which the clusters X_C around a "mother" point were modelled as independent Poisson processes with intensity functions

$$\lambda_C(\xi) = \alpha k(\xi - c; \omega) \exp(z_{1:p}(\xi) \beta_{1:p}^T)$$

where $\alpha > 0$, $\beta_{1:p}$ is a $1 \times p$ vector of regression parameters and k is a probability density depending on a parameter $\omega > 0$ that determines the spread of daughter points around the cluster centre c . The regression parameters describe the dependence of the spatial pattern on covariates such as soil or topographic variables.

Parameter estimation exploited the fact that second order summary statistics, in particular the K -functions are known in closed form such that minimum-contrast methods could be applied. The result indicated that there is a significant association of both species with the slope habitat but that they do not favour specific altitudes.

Further application and development of these techniques will undoubtedly yield major advances in our understanding of habitat relationships for tropical trees and other plant communities. For example using the 50-ha plot data from Pasoh, Peninsular Malaysia, recent analyses have shown that up to 60% of tree species had significant associations with one or more of ten soil chemical variables. See Figure 3 for a plot of the locations of the species *Aporosa bracteosa* on the estimated surface for calcium (Gimona *et al.*, unpublished data).

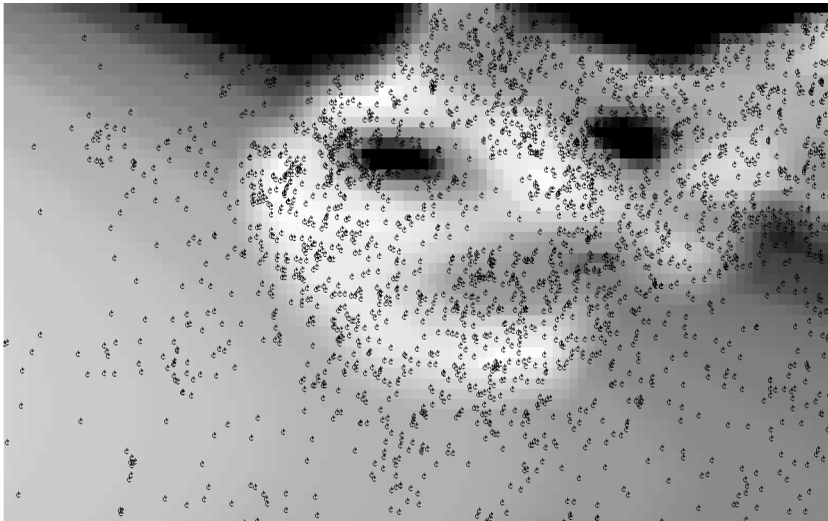


FIG 3. — Locations of species *Aporosa bracteosa* with estimated levels of Calcium.

4.3. Density dependent processes

The *Janzen-Connell hypothesis* (Janzen, 1970; Connell, 1971) predicts that the mortality of seeds and seedlings is positively related to conspecific seed or seedling density, and negatively related to increasing distance from a conspecific adult. According to this hypothesis, increasing distance of a seed or seedling from an adult plant, and/or lower seedling density, reduces susceptibility to natural enemies and increases its survival. This mechanism would result in a recruitment advantage for locally rare species and regulate the upper limit of species abundance.

The Janzen-Connell hypothesis makes very clear predictions about the spatial dependencies in the temporal dynamics of plant populations that are amenable to analysis using spatial point process models. For example, it predicts a less aggregated distribution of adult trees than juveniles. An appropriate model would seek to analyse the effects of habitat specialisation and density dependent interactions, as proposed by Diggle *et al.* (2006).

Exploratory data analysis based on second order summary statistics revealed that a large proportion of the species in the Australian shrubland vegetation exhibit clustering. The Janzen-Connell hypothesis predicts that species' survival is improved with increasing distance from the location of a conspecific and decreases with the number of conspecific individuals in close vicinity. This would result in older individuals tending to form a more regular pattern than juvenile individuals of the same species that might be clustered as a result of initial seed dispersal. However, the spatial patterning of juvenile and mature individuals cannot be compared in this community because of the unique age/size structure of many of its constituent species. Species that resprout are potentially hundreds of years old whereas others (seeder plants)

have been present for at most ten years- the period since the last fire. Comparing the spatial patterning between the species would result in a confounding of species and age effects.

The analysis of the Australian data set revealed that some resprouters, after having survived for extremely long times, still exhibit spatial clustering. This may provide evidence against the Janzen-Connell hypothesis for that community. However, spatial point process methods would contribute to the discussion in a more detailed way if data on the age of the individuals were available. More specifically, PCA methods, as described in Illian *et al.* (2005), may be applied to analysing the underlying structure in the variation in spatial patterns among species. For example, species might be divided into age classes and PCA methods can be used to analyse the differences in spatial patterns with regard to species and age class. If different age-groups of the same species are allocated to groups with different spatial behaviours, and most notably with different degrees of clustering, this may support the Janzen-Connell hypothesis.

Many detailed analyses of the tropical forest plot data have addressed the role of density dependence in regulating the abundance and distribution of the resident tree populations (Wills *et al.* 1997, 2006; Hubbell *et al.* 2001, Peters 2003). There is strong evidence from this recent work of pervasive negative and positive relationships between local conspecific adult density and recruitment and mortality rates respectively of saplings across multiple species. Only one of these papers considers the relationship of mortality and recruitment to local adult density in continuous space whilst allowing for the spatial autocorrelation in mortality, and here also the Janzen-Connell hypothesis is supported (Hubbell *et al.* 2001). Therefore, there is considerable scope for developing spatial point process models to conduct more realistic tests of the Janzen-Connell hypothesis taking account of additional factors such as spatial structure, environmental covariates, individual size, and phylogenetic relatedness, across multiple sites and species. These models would enhance significantly our understanding of the relative importance of different processes in the maintenance of species richness in tropical tree communities.

5. Discussion

This paper proposes the application of spatial point process methodology to biodiverse plant communities with the aim of gaining an improved understanding of the mechanisms that sustain biodiversity. Current studies have shown promising results but there is definitely scope for further work, most notably in the context of the rainforest data sets mentioned in the introduction.

In particular, current modelling approaches need to be extended further to incorporate simultaneously interactions within species, among species and between species and the environment. Further, multivariate data analysis methods for spatial point pattern data need to be extended as they are highly relevant for the multivariate data sets of biodiverse communities. This may for example include the application of principal component analysis but also

canonical correlation analysis to a data set of juvenile and adult (rainforest) trees.

6. Acknowledgements

This paper arose from discussions during the working group on spatial analysis of tropical forest biodiversity funded by UKPopNet and the Centre for Population Biology. We thank Richard Law for thoughtful discussions and comments on a previous version of the manuscript. We thank the Center for Tropical Forest Science for permission to use the data on *Aporosa bracteosa* from the Pasoh plot, which has been funded by the Forest Research Institute of Malaysia, U.S. National Science Foundation, Smithsonian Tropical Research Institute and the Center for Global Environmental Research at the National Institute for Environmental Studies, Japan.

References

- ADLER R. (1981). *The Geometry of Random Fields*. Wiley, New York.
- ARMSTRONG P. (1991). Species patterning in the heath vegetation of the Northern Sandplain. Honours thesis, University of Western Australia.
- BADDELEY A., J. MØLLER and R. WAAGEPETERSEN (2000). Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* 54, 329-350.
- BELL G. (2001). Neutral macroecology. *Science* 293, 2413-2418.
- CARDINALE B. J., A. IVES and P. INCHAUSTI (2004). Effects of species diversity on the primary productivity of ecosystems: extending our spatial and temporal scales of inference. *Oikos* 104, 437-450.
- CHAVE J. (2004). Neutral theory and community ecology. *Ecology Letters* 7, 241-253.
- CHESSON P. L. (2000). Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics* 31, 343-366.
- CONDIT R., P. S. ASHTON, P. BAKER, S. BUNYAVEJCHEWIN, S. GUNATILLEKE, N. GUNATILLEKE, S. HUBBELL, R. FOSTER, A. ITOH, J. LAFRANKIE, H. LEE, E. LOSOS, N. MANOKARAN, R. SUKUMAR, and T. YAMAKURA (2000). Spatial patterns in the distribution of tropical tree species. *Science* 288, 1414-1418.
- CONDIT, R., N. PITMAN, E. G. LEIGH Jr., J. CHAVE, J. TERBORGH, R. B. FOSTER, P. NUNEZ, S. AGUILAR, R. VALENCIA, G. VILLA, H. C. MÜLLER-LANDAU, E. LOSOS, and S. P. HUBBELL (2002). Beta diversity in tropical forest trees. *Science* 295, 666-669.
- CONNELL J. (1971). On the role of natural enemies in preventing competitive exclusion in some marine animals and in rainforest trees. In P. den Boer and G. Gradwell (Eds.), *Dynamics of Populations*, pp. 298-313. Centre for Agricultural Publishing and Documentation, Wageningen, The Netherlands.
- COOMES D. A., M. REES and L. TURNBULL (1999). Identifying aggregation and association in fully mapped spatial data. *Ecology* 80, 554-565.
- COX D. R. (1955). Some statistical models related with series of events. *Journal of the Royal Statistical Society Series B* 17, 883-904.

- CRAWLEY M. (1997). The structure of plant communities. In M. Crawley (Ed.), *Plant Ecology*, pp. 475-531. Blackwell Publishing, Oxford.
- DALEY D. J. and D. VERE-JONES (1988). *An Introduction to the theory of point patterns*. Springer-Verlag, New York.
- DE MAZANCOURT C. (2001). Consequences of community drift. *Science* 293, 1772.
- DEANGELIS D. L. and L. J. GROSS (1992). *Individual based Models and Approaches in Ecology: Populations, Communities and Ecosystems*. Chapman & Hall, London.
- DIECKMANN U., R. LAW and J. METZ (2000). *The Geometry of Ecological Interactions – Simplifying spatial complexity*. Cambridge University Press, Cambridge.
- DIGGLE P. (1983). *Statistical Analysis of Spatial Point Patterns*. Academic Press, London.
- DIGGLE P. (2003). *Statistical Analysis of Spatial Point Patterns*, 2nd ed. Hodder Arnold, London.
- DIGGLE P. J., V. GÓMEZ-RUBIO, P. E. BROWN, and A. G. CHETWYND (2006). Second-order analysis of inhomogeneous spatial point processes using case-control data. *Biometrics*.
- DIXON K. (2005). Personal communication.
- DUIVENVOORDEN J. F., J. C. SVENNING and S. J. WRIGHT (2002). Beta diversity in tropical forests. *Science* 295, 636-637.
- DURRETT R. and S. LEVIN (1998). Spatial aspects of interspecific competition. *Theoretical Population Biology* 53, 30-43.
- ELTON C. (1927). *Animal Ecology*. Sidgwick and Jackson, London.
- GASTON K. J. and S. L. CROWN (2005). Editorial: Neutrality and the niche. *Functional Ecology* 19, 1-6.
- GRINNELL J. (1917). The niche relationships of the Californian thrasher. *Auk* 34, 427-433.
- GRUBB P. J. (1977). The maintenance of species-richness in plant communities: the importance of the regeneration niche. *Biological Reviews* 52, 107-145.
- HARMS K. E., R. CONDIT, S. P. HUBBELL and R. B. FOSTER (2001). Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. *Journal of Ecology* 89, 947-959.
- HECTOR A., B. SCHMID, C. BEIERKUHNEIN, M. CALDEIRA, M. DIEMER, P. DIMITRAKOPOULOS, J. FINN, H. FREITAS, P. GILLER, J. GOOD, R. HARRIS, P. HOGBERG, K. H. DANELL, J. JOSHI, A. JUMPPONEN, C. KRNER, P. LEADLEY, M. LOREAU, A. MINNS, C. MULDER, G. O. G. S. OTWAY, J. PEREIRA, A. PRINZ, D. READ, M. S. LORENZEN, E. SCHULZE, A. SIAMANTZIOURAS, E. SPEHN, A. TERRY, A. TROUMBIS, F. WOODWARD, S. YACHI, and J. LAWTON (1999). Plant diversity and productivity experiments in European grasslands. *Science* 286, 1123-1127.
- HUBBELL S. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Monographs in Population Biology 32, Princeton University Press.
- HUBBELL S. P. (1979). Tree dispersion, abundance and diversity in a tropical dry forest. *Science* 203, 1299-1309.
- HUSTON M., D. L. DEANGELIS, and W. M. POST (1988). New computer model unify ecological theory. *BioScience* 38, 682-691.
- HUTCHINSON G. E. (1957). Concluding remarks. *Cold Spring Harbor Symposium on Quantitative Biology* 22, 415-457.

- HUTCHINSON G. E. (1959). Homage to santa rosalia, or why are there so many kinds of animals? *American Naturalist* 93, 145-159.
- ILLIAN J. B. (2006). Spatial point process modelling of a biodiverse plant community. PhD thesis, University of Abertay Dundee.
- ILLIAN J. B., E. BENSON, J. CRAWFORD, and H. J. STAINES (2004). Multivariate methods for spatial point processes – a simulation study. In A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan (Eds.), *Spatial point process modelling and its applications*, pp. 125-130. Castelló de la Plana: Publicacions de la Universitat Jaume I.
- ILLIAN J. B., E. BENSON, J. CRAWFORD, and H. STAINES (2006). Principal component analysis for spatial point patterns. In A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan (Eds.), *Case studies in spatial point process modelling*. Springer, New York.
- ILLIAN J. B., J. MØLLER, and R. P. WAAGEPETERSEN (2007). Spatial point process models for a complex plant community. To appear in *Journal of Ecological and Environmental Statistics*.
- JANZEN D. H. (1970). Herbivores and the number of tree species in tropical forests. *American Naturalist* 104, 501-528.
- JUDSON O. (1994). The rise of the individual-based model in ecology. *TREE* 9, 9-14.
- LAW R., D. MURRELL and U. DIECKMANN (2003). Population growth in space and time: spatial logistic equations. *Ecology* 84, 252-262.
- LOREAU M. (2000). Biodiversity and ecosystem functioning: recent theoretical advances. *Oikos* 91, 3-17.
- LOREAU M., S. NAEEM, P. INCHAUSTI, J. BENGTSSON, J. P. GRIME, A. HECTOR, D. U. HOOPER, M. A. HUSTON, D. RAFFAELLI, B. SCHMID, D. TILMAN, and D. A. WARDLE (2001). Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* 294, 804-808.
- MATEU J., J. L. Us'o, and F. Montes (1998). The spatial pattern of a forest ecosystem. *Ecological Modelling* 108, 163-174.
- MØLLER J. and R. WAAGEPETERSEN (2003a). An introduction to simulation-based inference for spatial point processes. In J. Møller (Ed.), *Lecture Notes in Statistics* 137, pp. 143-198. Springer-Verlag, New York.
- MØLLER J. and R. WAAGEPETERSEN (2003b). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall/CRC, Boca Raton.
- MOUQUET N., J. L. MOORE, and M. LOREAU (2002). Plant species richness and community productivity: why the mechanism that promotes coexistence matters. *Ecology Letters* 5, 56-65.
- MURRELL D. J., D. W. PURVES, and R. LAW (2001). Uniting pattern and process in plant ecology. *Trends in Ecology and Evolution* 16, 529-530.
- PURVES D. and R. LAW (2003). Heteromyopia and the spatial coexistence of similar competitors. *Ecology letters* 6, 48-59.
- PURVES D. W. and S. W. PACALA (2005). Ecological drift in niche-structured communities: neutral pattern does not imply neutral process. In D. Burslem, M. Pinard, and S. Hartley (Eds.), *Biotic Interactions in the Tropics*, pp. 107-138. Cambridge University Press, Cambridge.
- RAMSAY J. and B. SILVERMAN (1997). *Functional data analysis*. Springer, New York.
- RAMSAY J. and B. SILVERMAN (2002). *Applied functional data analysis*. Springer, New York.

- REGAN H., R. LUPIA, and M. BURGMANN (2001). The currency and tempo of extinction. *American Naturalist* 157, 1-10.
- RIPLEY B. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability* 13, 255-266.
- SCHLATHER M. (2001). On the second-order characteristics of marked point patterns. *Bernoulli* 7, 99-117.
- SCHOENER T. W. (1989). The ecological niche. In J. Cherrett (Ed.), *Ecological Concepts*, pp. 79-113. Blackwell Scientific Publications, Oxford.
- STOLL P. and J. WEINER (2000). A neighbourhood view of interactions among individual plants. In U. Dieckmann, R. Law, and J. Metz (Eds.), *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*, pp. 11-27. Cambridge University Press, Cambridge.
- STOYAN D., W. KENDALL, and J. MECKE (1995). *Stochastic Geometry and its Applications* (2nd ed.). John Wiley & Sons, London.
- STOYAN D. and A. PENTTINEN (2000). Recent applications of point process methods in forestry statistics. *Statistical Science* 1, 61-78.
- STOYAN D. and H. STOYAN (1994). *Fractals, random Shapes and Point Fields*. John Wiley & Sons, London.
- TILMAN D. (1994). Competition and biodiversity in spatially structured habitats. *Ecology* 75, 2-16.
- TILMAN D., P. B. REICH, J. KNOPS, D. WEDIN, T. MIELKE, and C. LEHMAN (2001). Diversity and productivity in a long-term grassland experiment. *Science* 294, 843-845.
- TILMAN D., D. WEDIN, and J. KNOPS (1996). Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* 379, 718-720.
- TURKINGTON R. A. and J. L. Harper (1979). The growth, distribution and neighbour relationships of *Trifolium repens* in a permanent pasture. I ordination, pattern and contact. *Journal of Ecology* 67, 201-208.
- VAN LIESHOUT M. (2000). *Markov point processes and their applications*. Imperial College Press, London.