CrossMark

ORIGINAL PAPER

Improving the usability of spatial point process methodology: an interdisciplinary dialogue between statistics and ecology

Janine B. Illian[1] · **David F. R. P. Burslem**[2]

© The Author(s) 2017. This article is an open access publication

Abstract The last few decades have seen an increasing interest and strong development in spatial point process methodology, and associated software that facilitates model fitting has become available. A lot of this progress has made these approaches more accessible to users, through freely available software. However, in the ecological user community the methodology has only been slowly picked up despite its obvious relevance to the field. This paper reflects on this development, highlighting mutual benefits of interdisciplinary dialogue for both statistics and ecology. We detail the contribution point process methodology has made to research on biodiversity theory as a result of this dialogue and reflect on reasons for the slow take-up of the methodology. This primarily concerns the current lack of consideration of the usability of the approaches, which we discuss in detail, presenting current discussions as well as indicating future directions.

1 Introduction

1.1 Point process models in the modern world

In the past, complex statistical methods were not directly accessible to the applied user. Data analyses that required statistical approaches beyond those covered in introductory statistics textbooks would be done using software tailor-made for the specific applica-

✉ Janine B. Illian
jbi@st-andrews.ac.uk

David F. R. P. Burslem
d.burslem@abdn.ac.uk

1 School of Mathematics and Statistics, University of Saint Andrews, St Andrews, UK

2 School of Biological Sciences, University of Aberdeen, Aberdeen, UK

🖄 Springer

tion by an expert. Nowadays, freely available, sophisticated software packages such as R are in common use, and at the same time—in these days of Big Data—increasing amounts of data are collected (Tippmann 2015; Decan et al. 2015) and new journals have emerged to communicate advances in statistical methods to non-statistical audiences. As a result, users have both a stronger need for data analysis and an increasing awareness of the existence of the advanced methodology since it is now less "hidden" from them in inaccessible statistical journals. Hence, in ecology, in particular, it is now relatively common for non-specialists to use advanced statistical methodology and associated software without direct support from a statistician (Zuur et al. 2010).

Some might argue that this is a retrograde step and to avoid misuse, complex statistical methods should only be used by experts. However, this argument has been largely pre-empted by the high demand for new statistical methods and their widespread uptake among user communities—as a result researchers outside statistics are already using complex methods, and higher education institutions are expanding their provision of statistical training for undergraduate and postgraduate biologists. Rather than worrying about the misuse of methods, we argue that statisticians should consider the usability of methodology an important part of method development. The issue is not misuse of methodology or the ignorance of the user community—we cannot possibly expect them to become experts as well, considering how long it has taken each one of us to get where we are. The issue is rather that methods have often been developed from the point of view of a statistician, and little thought has gone into their relevance and the practicality of applying them. We outline in this paper how we believe that it is possible to improve the "usability" of methodology, i.e. its suitability for use by non-experts, by considering the *relevance* and *practicability* of statistical methodology through an ongoing interdisciplinary dialogue.

Communicating practical advice in an accessible way is an important component of statistical method development. More specifically, while it is now relatively common to publish an R-library associated with a specific statistical development, we still need to appreciate that making statistical methods accessible to the research community involves not only the development of methodology and providing the software but also considering the practicalities of using it and communicating these details in a style that is accessible to intended users. Making model fitting feasible by developing computationally efficient approaches to reduce running times is clearly a step in the right direction (Pélissier and Goreaud 2015). However, enhancing the practicality of modelling requires not only improvements to model fitting. The practicality of the entire statistical analysis from model construction via prior choice to interpretation is equally relevant for users with real data sets and real scientific questions. Similarly, the increasing complexity of the methodology not only makes model fitting more complex, but also implies that all steps in a statistical analysis become more complex. An effective impact on the practicalities of statistical analysis hence requires an understanding of what the users need and what they are interested in, which in turn calls for statisticians to interact with the data as well as with the general research questions. This generates a need for an interdisciplinary approach to research which integrates statisticians with users during all stages of the research programme.

In this paper, we illustrate the importance of interdisciplinary dialogue among statisticians and ecologists using past and current examples of the application of spa-

tial point process methodology and how it has contributed to topical debates in both ecology and statistics. This discussion is particularly relevant for three reasons. Firstly, relative to other statistical methods used by ecologists, such as distance sampling or capture–re-capture methods, the relevance of spatial point process methodology to ecological users is less obvious. Secondly, unlike other general statistical methods such as generalised linear models and related approaches, spatial point process methodology has long been seen as a rather niche sub-area of statistics that is of theoretical interest to some, with few relevant data sets available and hence of little practical relevance. Thirdly, due to their complexity, spatial point process models have been historically particularly hard to fit, making them difficult to use in relevant applications and particularly inaccessible to applied users. As a result, neither applied statisticians nor statistically interested ecologists, which represent the community who would be inclined to engage in an interdisciplinary dialogue, have been well acquainted with the methodology, and hence, an interdisciplinary exchange has lagged behind that in other areas of (applied) statistics or statistical ecology. The slow uptake of spatial point processes by ecologists has occurred despite their suitability for analysis of the patterns formed by individuals in space—the very objects that ecologists are interested in.

1.2 Relevance of spatial point processes to ecology

The field of ecology addresses the interactions among organisms and their biotic and abiotic environments. Since individuals exist in space and time, local spatial structure of the physical environment and biotic neighbourhood impacts ecological processes but at the same time these processes help to generate and alter spatial structure in an ecological community. Therefore, the spatial structure may partially represent underlying processes, although the extent to which these processes may be uncovered by analysis has been a subject of debate in ecology for many years (Law et al. 2009). The obvious practical implication should be that analysing the spatial pattern of an ecological community may help the understanding of the underlying ecological processes (Brown et al. 2015).

However, in the past, data on individuals in space and time were rarely available, so statistical methods as well as ecological research dealt with data that were either not spatially explicit or that were aggregated into spatial units such as grid cells. Technical advances have made it easier to collect data on the spatial locations of individuals, and spatially explicit data sets have become commonly available. This enables us to consider the locations of individuals directly, rather than aggregating them in space and hence ignoring the spatial structure and information contained in it (Anderson-Teixeira et al. 2015). In other words, the data that reflect spatial structure in ecosystems are patterns formed by the locations of individuals in space, i.e. spatial point patterns. The statistical representation of these is a *spatial point process*, which models the spatial location of objects—here individuals or groups of individuals—in space. However, despite often being motivated as models of spatial patterns observed in nature—specifically of trees—publications on spatial point process modelling mainly resided within the statistical literature until rather recently, with only limited applications outside the statistical literature that directly concern or answer ecological (or

other scientific) questions. This has changed over the last decade or so, partly due to the improved accessibility of suitable software as well as an improved dialogue between statisticians and ecologists interested in the respective fields. However, spatial point processes are still not considered part of the standard toolbox of statistical methods used by ecologists and the methods are only now being picked up by the ecological community in general—despite ecologists being highly numerate and keen on using appropriate statistical methodology in general (Law et al. 2009).

   This paper aims to reflect on the benefits of interdisciplinary work in the light of the ongoing dialogue between ecology and spatial point process statistics. It illustrates how the dialogue has contributed to both areas and highlighted weaknesses in relevant statistical approaches—in the specific context of ecology as well as in applications in general. We initially provide a brief summary of general point process methodology for those readers unfamiliar with the particular literature (Sect. 2). We then outline how interdisciplinary dialogue has improved the *relevance* of point process methodology using the example of a specific ecological debate that point process methodology has contributed to in recent years, that of opposing theories of species coexistence in ecological communities (Sect. 3). Section 4 discusses how interdisciplinary dialogue has lead to the development of complex statistical methods that are relevant in other areas of application. Section 5 discusses the importance of considering the *practicability* of the methodology in the context of complex point process models and leads on to a general discussion of what aspects of point process methodology are currently being discussed as a result of an ongoing dialogue between ecologists and statisticians (Sect. 7).

2 Spatial point processes in a nutshell

This section provides a brief background to spatial point processes; for a more formal definition and technical details, refer to the relevant textbooks including Lieshout (2000), Møller and Waagepetersen (2004), Diggle (2003), Illian et al. (2008), Wiegand and Moloney (2013) and Baddeley et al. (2015); notation is as in Illian et al. (2008). Formally, a spatial point process is a random variable N that is defined on subsets B of $\mathbb{R}^d$, and $N(B)$ is the random number of points in B. Realisations of N are called *point patterns*. For $x \in \mathbb{R}^2$, consider the expected number of points in B and a density function $\lambda(.)$, referred to as the *intensity function*:

$$\mathbb{E}(N(B)) = \int_B \lambda(x)\mathrm{d}x,$$

which plays a central role in spatial point process models as we will see below.

   Point patterns where information on the properties of the points is available may be described by a similar random variable, a *marked point process M*. For a subset B of $\mathbb{R}^d$ and a subset C of $\mathbb{R}^1$, $M(B \times C)$ denotes the random number of marked points $[x; m(x)]$ with $x \in B$ and $m(x) \in C$. The marks can be qualitative, leading to a multi-type point process, or quantitative and may be included in a statistical analysis of a point pattern.

Note that the point process literature usually assumes one of two scenarios, referred to as "Window sampling" and the "Small world Model" in Baddeley et al. (2015). Both scenarios imply that the entire area relevant in a study has been surveyed and that either all individuals represented by the points in a pattern have been detected, or the proportion of non-detected individuals is known and the same everywhere in space. We will discuss issues with this assumption in Sect. 6.

Generally speaking, point patterns may be analysed just like any data type; we consider a) exploratory methods, i.e. summary characteristics that describe the properties of a point pattern, and b) point process models that relate the structure of the pattern to other variables, for example spatial covariates characterising the local environment. Model fitting is often informed by an initial exploratory analysis of a point pattern to suggest a specific type of point process model. We briefly sketch some of the main exploratory and model fitting approaches here.

2.1 Summary characteristics

A number of summary characteristics have been discussed in the literature, including several indices that summarise properties of the spatial pattern in a single number. In practice, the most versatile characteristics are summary characteristics that describe a point pattern through a function of distance and there is a large literature on the estimation and interpretation of these functions. Most prominent are second-order summary characteristics such as the K-function, its variant the L-function and the pair correlation function. They consider the occurrence of points for a range of relevant distances by considering the average number of other points found within (K-function and L-function) or at (pair correlation function) distance r from a typical point in a pattern. Related functional summary characteristics have been developed for marked point patterns for both qualitative and quantitative marks to establish relationships between the point pattern and the marks. For instance, the cross-pair correlation function considers the co-occurrence of individuals of different types of points and the mark correlation function considers if points at a specific distance have particularly small (or large) marks.

Classically, the empirical functions are compared to functions analytically known for appropriate null models such as the most basic relevant null model, the Poisson process, which reflects complete spatial randomness (CSR). Its interpretation involves considering the nature of the deviation from the functions for those null models. For example, the K-function for the Poisson process has the following form: $K(r) = \pi r^2$ for $r \geq 0$ ($L(r) = r$ for $r \geq 0$ for the L-function and $g(r) = 1$ for $r \geq 0$ for the pair correlation function). For an observed point pattern, the summary characteristics have to be estimated and the literature discusses suitable estimators. Similarly, suitable approaches to edge correction have been considered that avoid estimation bias resulting from censoring outside the observation window. It is difficult to determine when edge corrections are necessary as both the number of points and the structure of the pattern impact on the susceptibility of an analysis to a bias caused by edge effects. In general, smaller patterns are more likely to be affected by edge effects as they contain less information on spatial structure than

larger patterns and are hence more sensitive to a bias. Similarly, strongly clustered or strongly inhomogeneous patterns are more likely to be affected by an edge effect. In an extreme case, one might observe a few points close to the edge that form part of a large (unobserved) cluster and that are relatively far away from other clusters. This would lead to an impression that these points have few other points in their vicinity and underestimate the strength of clustering.

In an analysis, the estimated empirical functions are then compared to those for the relevant null model. For instance, if an empirical function takes on values above/below what is expected for the Poisson process at specific distances, this is interpreted as attraction/repulsion among points at these distances. Considering null models for marked point patterns becomes more complicated due to the increased number of relevant dependencies that might be of interest, see discussions in Goreaud and Pélissier (2003) and Illian et al. (2008) for details. In order to aid the interpretation of the estimated summary characteristics for an empirical pattern, simulation envelopes based on repeated simulation of the relevant null model are used as informal tests, while formal tests are used to assess the significance of the deviation (Myllymäki et al. 2015).

2.2 Point process models

The homogeneous Poisson process with constant intensity function $\lambda(x) = \lambda$ is the appropriate model of complete spatial randomness and useful as a null model in many contexts. In practice, however, patterns are not random and null models that reflect different spatial patterning structures are often more relevant (Wiegand and Moloney 2004). In addition, just as any statistical model, spatial point processes may also be used as analysis tools directly. Hence, Poisson processes have been generalised to construct more complex and more flexible models that reflect more realistic, yet more complex structures in point patterns. These include Neyman–Scott processes, which are models of clustered patterns, where the clusters are generated by a (Poisson) "parent process" and the "children" are distributed around these parents following some two-dimensional distribution. The class of Gibbs (or Markov) processes is a larger class of point process models that is highly relevant in practice. These explicitly include terms reflecting local interaction among the objects represented by the points (Lieshout 2000). The class of Cox processes consists of doubly stochastic point processes in which the intensity function is a random field. In the special case of a *log Gaussian Cox process* (Møller et al. 1998), the intensity surface is modelled as

$$\log(\Lambda(s)) = Z(s),$$

where $Z(s)$ is a Gaussian random field. The random field can be constructed in various ways, e.g. as a linear combination of covariates and a spatially structured (Gaussian) error field, making these models very flexible. This flexibility has been exploited recently, and Sects. 4 and 5 discuss specific examples that form the background to the main topic of this paper.

In general, fitting point process models is not straightforward since for any but the simplest models the likelihood is often analytically intractable. Various approaches

have been considered in both a frequentist and a Bayesian setting, using explicit approximations or pseudolikelihood approaches, and MCMC or integrated nested Laplace approximation (INLA), respectively (Ogata and Tanemura 1985; Geyer and Møller 1994; Baddeley and Turner 2000; Møller and Waagepetersen 2004; Rue et al. 2009; Illian et al. 2012b).

Within the ecological literature, point process models have rarely been fitted to data but descriptive methods have been used more frequently (Haase 1995; Wiegand and Moloney 2004; Illian et al. 2008; Law et al. 2009; Brown et al. 2011), but see Cornulier and Bretagnolle (2006) and Wiegand et al. (2007). We will illustrate below how the use of variants of the classical summary characteristics discussed here has contributed to an improved understanding of ecological theory and how ecological data sets have inspired the construction of more complex point process models, in particular marked point process models and point processes with complex observation processes.

3 Making spatial point process methodology relevant—the spatial footprint of species coexistence

3.1 Biodiversity—coexistence in space

Biodiversity is declining as a result of anthropogenic impact on the world's ecosystems, and both conservationists and politicians have been discussing preventative measures and targets for reduction of biodiversity loss (Brooks et al. 2006; Rands et al. 2010; Tittensor et al. 2014). A lot of work has gone into documenting and measuring the decline—not least in order to be able to assess whether preventative conservation measures are effective. As a result, and due to the complexity of the concept, measuring biodiversity itself and global biodiversity indicators have been the topic of many publications and a subject of statistical ecology research (Chape et al. 2005; Magurran 1988; Magurran and McGill 2010; Studeny et al. 2010; Jones et al. 2011). However, since it is not sufficient to only measure and document biodiversity and biodiversity change, current research attempts to understand the role of biodiversity in ecosystem functioning, the potential consequences of the loss of biodiversity on ecosystem functioning, which is, ultimately, the motivation for species conservation (Ferraro and Pattanayak 2006; Schmid et al. 2009; Naeem et al. 2009). It is hence of primary interest to ecologists to understand the processes that allow large numbers of species to coexist in biodiversity hotspots, such as tropical rainforests (Wright 2002). These rainforests boast a particularly high level of biodiversity where one hectare of tropical rainforest can contain 600 or more tree species (Slik et al. 2015). However, processes that generate and maintain this high diversity are only poorly understood despite a proliferation of theories (Wright 2002).

Understanding how biodiversity is maintained implies understanding the coexistence of a large number of different species, and mechanisms of coexistence have been studied in ecology for a long time. Darwin himself suggested that communities were structured through competition, which resulted in the competitive exclusion principle (Pocheville 2015) and the formulation of *niche theory*, which implies that losing species does change community dynamics and hence that a decline in biodiver-

sity impacts the functioning of an ecosystem. Most prominently, the so-called *neutral theory* postulates that this is not the case and that individuals are competitively equivalent (Hubbell 2001), concluding that the diversity of communities is maintained by a balance between immigration and extinction that is independent of species identities. Other mechanisms that have been suggested as relevant for maintaining species coexistence and hence biodiversity are *heteromyopia*, which postulates that the spatial scale of the interactions among individuals from different species differs from that among individuals from the same species (Murrell and Law 2003; Purves and Law 2003) and the *Janzen-Connell hypothesis* (Janzen 1970; Connell 1971), which postulates a competitive advantage for individuals of rare species or those located at increasing distances from conspecifics that might act as a reservoir of host-specific natural enemies.

All these concepts are either explicitly or implicitly spatial, implying that coexistence of species is directly linked to the nature of the interactions of individuals with their local environment, in terms of both competition with other individuals and their relationship with abiotic properties of the environment (Law et al. 2000; Illian and Burslem 2007). Hence, considering the local spatial structure rather than averaging across space better reflects the processes relevant to species coexistence. However, while ecologists have realised this when simulating communities using individual-based models, when describing communities and comparing different simulation approaches they have traditionally used summary statistics that relate to first order properties rather than those that characterise the spatial behaviour of individuals. As a result, approaches that tried to distinguish between different theories of species coexistence based on first order summary characteristics, have failed to do so (Chave 2004; McGill et al. 2007; Zillio and Condit 2007). Recent work has demonstrated that the second-order spatial structure of communities provides greater insights into the mechanisms maintaining species coexistence (Brown et al. 2011, 2013; Baldeck et al. 2013), and we outline below how interdisciplinary dialogue has contributed to this result.

3.2 Distinguishing theories—using summary characteristics

In order to describe the structure and nature of an ecological community, it is necessary to find a descriptor that summarises the properties of that community. Commonly, the species abundance distribution (SAD), i.e. the abundance of individuals for each different species in an ecological community plotted against the rank in abundance, has been used for the purpose. It is well known that the SAD takes on a characteristic shape reflecting the presence of a few very abundant species and many that are rare, which is a pattern that has been documented across multiple communities of different kinds of organisms (Hubbell 2001; McGill et al. 2007). Due to their universality, SADs have often been used to support the validity of neutral theory, as communities simulated based on this theory also result in a SAD with the classical shape (Hubbell 1979, 2001).

Keeping in mind that species' relative abundances are averaged over space to calculate a SAD, resulting in a clear loss of information contained in an ecological

community that has been described using spatially explicit data, they simply might not be a suitable summary characteristic when it comes to distinguishing predictions derived from theories on the structure of a community. Indeed, either explicitly—such as the Janzen–Connell hypothesis —or implicitly, theories on the processes that maintain species coexistence make a prediction about the spatial structure of a community, suggesting that spatial properties of populations within communities may be used to assess the validity of these theories and reveal differences among theories based on their predictions.

With strong reference to Brown et al. (2011, 2013), we detail here how this idea has been assessed based on two main alternative approaches: (a) simulating artificial communities over many generations according to postulated mechanisms as predicted from various theories of biodiversity followed by an analysis of the distinct spatial signals that emerge at equilibrium and (b) describing the spatial pattern of natural highly biodiverse communities to determine whether these patterns match the spatial predictions of alternative theories of species coexistence.

3.2.1 Patterns in simulated communities

In this context, spatially explicit simulations consider the behaviour of individuals in space, relative to other individuals, typically through stochastic individual-based models (IBMs); full details of the implementation of these methods can be found for example in Law and Dieckmann (2000). In particular, Brown et al. (2011) used IBMs to generate artificial communities according to the four different mechanisms discussed above, niche theory, neutral theory, heteromyopia and the Janzen-Connell hypothesis and then assessed the spatial signal in these communities. No attempts were made to parametrise these IBMs to generate simulations of a specific study system, but the interest was purely in the spatial signal resulting from the different theoretical constructs. Specifically, in Brown et al. (2011) individuals of different types (species) were generated on a torus (to avoid edge effects) of unit area in continuous time, where existing individuals can generate offspring in a specific location $x = (x_1, x_2)$ following an intrinsic probability per unit time of producing an offspring, b. This was multiplied by a bivariate Gaussian distribution with variance σ_b, so that the probability of a birth in a specific location decreases with distance from the individual "parent". The death of an individual of species i was determined based on an intrinsic death rate d_i. To mimic competition for resources, and hence density dependence, an interaction term was added to the intrinsic death rate. This term increased with increasing numbers of individuals in an individual's vicinity, where the influence of a neighbour of species j decreased with its distance from the target individual according to a bivariate Gaussian distribution with variance $\sigma_{d_{ij}}$. The interaction term was weighted by a weight d_{ij} to make the strength of the interaction dependent on species identities.

In the IBM simulating a *neutral model*, b, d, σ_b and $\sigma_{d_{ij}}$ were the same for all individuals and species, and constant over space. The IBMs for the three other postulated mechanisms were generalisations of that for the neutral model, where generalisation was achieved through varying how the probability of a death per unit time was determined. In the *niche model*, the death rate varied in space depending on a (simulated)

spatially varying covariate, mimicking spatially varying environmental conditions (see Brown et al. 2011 for details on the simulation of the environmental variable). In addition, different species were randomly allocated to be associated with a specific value of this environmental variable, mimicking niche preferences. To mimic *heteromyopia*, $\sigma_{d_{ij}}$ was chosen to be larger if $i = j$ than if $i \neq j$, whereas for the *Janzen–Connell effect* the interaction term was given a higher weight d_{ij} if $i = j$ than if $i \neq j$ to mimic the negative effects of conspecific neighbours which may arise through competition or shared natural enemies.

In order to compare simulations based on different theoretical assumptions, the co-occurrence of individuals from different species was assessed based on a pairwise summary characteristic, the cross-pair correlation function $g_{ij}(r), r = 0, \ldots, R$ (Diggle 2003; Illian et al. 2008). For each species pair i, j, this was aggregated as

$$A_{i,j} = \int\limits_0^R \log(g_{ij}(r)) dr.$$

The results of this study showed that SADs produced by models simulating communities were as variable within as among simulations generated under the assumptions of different species coexistence mechanisms. This clearly emphasised that SADs are not an appropriate tool to distinguish the various mechanisms. However, the spatial information as reflected in the distributions of the $A_{i,j}$ across species proved to be clearly differentiating among mechanisms. The most dramatic differences were found in the spread of these distributions, in particular between the patterns generated under the assumptions of the neutral and niche assembly mechanisms (Brown et al. 2011). Further work has looked into the use of other spatial summary characteristics in the same context and also considers the different mechanisms in combination rather than just in isolation, which is likely to be more realistic in natural systems (Brown et al. 2015). This research suggests that the spatial signals of niche and neutral models remain distinct even when their assumptions are implemented alongside those of alternative species coexistence mechanisms.

3.2.2 Applications to real ecosystems

The discrimination between different theoretical explanations of species coexistence using simulations is a useful exercise, but it is also important to investigate real communities and analyse spatial signals contained in these, as done, for example, in Condit et al. (2002), Seidler and Plotkin (2006), Debski et al. (2000, 2002) and Bagchi et al. (2011). Analysing the spatial signal in real communities, in particular those that possess high diversity, allows us to assess whether suggested mechanisms such as those discussed in the previous section are likely to maintain the diversity in these systems or not. To discuss this in detail, we need to return to niche theory and the associated competitive exclusion principle. It is well known that many species have strong habitat associations, which reflect a balance between differential habitat filtering and competition within plant neighbourhoods. Such species–habitat associations have been detected multiple times in tropical tree communities (e.g. Harms et al.

2001; Debski et al. 2002; Gunatilleke et al. 2006; Bagchi et al. 2011). However, in order to infer that these patterns are determined by habitat niche partitioning, it is necessary to uncouple the spatial signal of the environment from the pervasive spatial structuring associated with seed dispersal limitation (Hamill and Wright 1986). In real systems and for an individual site, clustering resulting from habitat association cannot be easily distinguished from the (potentially smaller scale) clustering resulting from local seed dispersal, unless strong assumptions are made or knowledge is available on the specific differences of the spatial scales at which these mechanisms operate. Hence, in many contexts, it is difficult to draw conclusions on their relative roles and the relative strength of their contributions to observed spatial patterns.

However, by considering a number of sites with varying degrees of environmental heterogeneity, and comparing spatial signals across sites Brown et al. (2013) were able to circumvent this problem. The study took advantage of the availability of data from numerous large plots across the world through a network coordinated by the Center for Tropical Forest Science (CTFS) (Losos and Leigh 2004) to compare the vast amounts of spatial information contained in these data sets. The results show that pairwise associations between species were more variable in sites displaying greater topographic heterogeneity, which is consistent with the hypothesis that topographic niches contribute to species' spatial structure irrespective of effects of seed dispersal limitation. Neutral theory cannot explain this relationship.

In summary, this recent work has shown that crucial ecological information is contained in the spatial structure of an ecological community and that this has previously been largely ignored in the ecological literature. Input from spatial point pattern statistics and a strong dialogue has provided a different view on a topical question and highlighted an issue with previous approaches. The approach that was taken in this context has been based on second-order summary characteristics of point patterns. However, there is a subtle but relevant difference in how this was done here as opposed to common approaches in the statistical literature—in the examples reviewed here the shape of the summary characteristics was compared under different scenarios, while common statistical practice is to compare the empirical result to a theoretical null model. While this contrast in approach is not overly statistically exciting as such, it provides a good example of a case where the standard approach that is adopted and relevant in the context of statistics, i.e. the comparison to a null model, provides little information of interest since it is known that populations and communities often show clustered patterns. The most obvious model to compare an empirical pattern to from the mathematical point of view does not necessarily reflect what is relevant for a particular scientific question. From the point of view of the usability and relevance of the methodology, it is important that users are made aware that it is appropriate to adapt the approach to suit the underlying question, e.g. that a summary characteristics cannot only be used to assess whether a specific pattern exhibits complete spatial randomness but also to distinguish between different patterns. Inferring process from patterns in plant populations and communities has a long history in ecology (Watt 1947; Harper 1977), and application of new spatial point process methods has the potential to improve both the description of patterns and the inference of mechanism, especially when coupled with experimental techniques.

4 Making spatial point process methodology relevant—complex marked point process models

In the previous section, we have emphasised that considering spatial structure is relevant to a better understanding of community dynamics in ecology and attempted to show how interdisciplinary dialogue has contributed to an increased awareness of suitable statistical methodologies for handling spatially explicit data. However, it is not enough to simply establish that existing methods are relevant but also to improve relevance by developing methodology that meets the needs of the user community. We will illustrate how dialogue has improved relevance by encouraging the development of new point process models and methods.

As mentioned in Sect. 1, a major barrier to the usability of point process models used to be that model fitting was complicated and computationally expensive and, as a result, little standard software was available until the seminal work by Baddeley and Turner (2000). They showed that, based on the pseudolikelihood, Gibbs processes can be fitted using standard software for generalised linear models. This led to the development of the R-library spatstat (Baddeley and Turner 2005; Baddeley et al. 2015), a very versatile and large library for point processes, whose functionality has now developed far beyond the fitting of Gibbs processes. spatstat provides a wide range of point process tools, many of which are likely to be relevant to applied users, and it has greatly improved the accessibility of point process methodology. However, while Gibbs processes have been applied in a small number of papers in an ecological context (DeSoto et al. 2010; Cornulier and Bretagnolle 2006), the development of the library has been strongly focused on implementing methods and fitting approaches for existing published summary characteristics and models rather than developing methods that have been motivated by concrete data structures or scientific studies. Despite their obvious potential relevance in an ecological context as models of inter-individual interactions, Gibbs processes have not become part of the statistical toolbox used by ecologists.

Another approach that has facilitated the fitting of point process models considers log Gaussian Cox processes (see Sect. 2.2). This class of models can be regarded as a special case of latent Gaussian models, where the latent Gaussian structure is the random intensity function. Latent Gaussian models and hence also log Gaussian Cox processes may be fitted with integrated nested Laplace approximation (INLA) in a Bayesian context (Rue et al. 2009). For this purpose, the continuous Gaussian random field is approximated by a Gauss Markov random field (GMRF) with local dependence structure. INLA speeds up model fitting and makes log Gaussian Cox processes more versatile, as these can now be fitted more quickly than previously, allowing the development of more complex models that are still realistically feasible to fit. In addition, the R libraries R-INLA and the emergent inlabru software (Lindgren and Rue 2015; Inlabru Development Team 2017) make fitting these models accessible to users as the syntax for model fitting is akin to that familiar to users, i.e. that used in R-functions such as lm or glm. Unlike spatstat, which was specifically developed for the analysis of point patterns, these packages may be used to fit a large range of general spatial and spatio-temporal (and other) models, including some particularly flexible point process models, and do not specifically pro-

vide functionality for analysing point patterns with summary characteristics. Again, by providing relevant, computationally efficient software, this development has made the methodology accessible to non-specialist users by making relevant, computationally efficient software available. But clearly software is not enough, as evidenced by the rather limited number of ecological publications using the approach to date. It is hoped that through effective interaction with the user community we will be able to make methodology more relevant—and providing usable software is just one step in the right direction. In this section, we outline how the construction of complex log Gaussian Cox processes has made spatial point process modelling more relevant in practice, by providing answers to ecological questions through direct interaction with ecologists rather than being mainly statistically interesting. However, we will also outline further below how this very interaction has challenged current approaches and practices and shown us that we still have a long way to go (Sect. 5).

Marked point processes have been considered in the statistical literature to a much lesser extent than unmarked point processes. This is perhaps because whilst formulating a stochastic process that characterises the spatial structure of a point pattern is mathematically particularly challenging, and hence interesting to scientists, additionally considering a mark is less so. This might be paired with a lack of relevant data sets being available to statisticians and hence the problem seeming to be less relevant. Method development in this context has mainly focused on summary characteristics for both quantitatively and qualitatively marked point patterns but has rarely considered marked point process models (Schlather 2001). This is particularly true for point processes with quantitative marks; multi-type point patterns have been discussed, for example, in Högmander and Särkkä (1999) and Illian et al. (2004). Specifically, in the literature, point processes where marks are dependent on the spatial pattern have rarely been constructed. Methods for testing for dependence between point patterns and marks have been discussed, but models have rarely been fitted to data and, if so, have often assumed independence between marks and the intensity of the pattern (Myllymäki and Penttinen 2009). However, when considering spatial structure in practice, and in particular in an ecological context, marks, i.e. the properties of the individuals, might be just as important as their location, whereas the ecological insight that may be gained from only considering the pattern formed by individuals is likely to be limited. In applications, the main interest may in fact be on the marks rather than the spatial pattern, while the spatial component may be included in a model to account for an underlying spatial structure that is likely to affect the pattern—perhaps even trivially so (Ledo et al. 2016). However, assuming that these properties do not depend on the spatial properties of the patterns appears rather unrealistic.

The flexibility of Cox processes resulting from the random intensity function coupled with access to fast model fitting software through R-INLA has made it feasible to develop realistically complex marked point processes that allow for dependence between marks and points and that can still be fitted within a reasonable time. The approach models the marks and the point pattern in a joint model with two or more likelihoods, which are linked as they share some common elements, in particular joined random fields (Illian et al. 2012b).

In R-INLA, the approximation of the continuous random field $\log(\Lambda(s)) = Z(s)$ through a GMRF can be done in two different ways. In the simplest case, the observa-

tion window is discretised into regular grid cells and the number of observed points in each cell is assumed to follow a Poisson distribution, given the latent field. However, the aggregation to counts in grid cells constitutes a clear loss of information and the choice of the size of the grid cells is often arbitrary. In addition, the regular grid structure is rather rigid and can be computationally wasteful. To address this, an approximation based on stochastic partial differential equations (referred to as the *SPDE approach* as introduced in Lindgren et al. (2011) and adapted for point processes in Simpson et al. (2016) is available in R-INLA. This approach provides a spatially continuous approximation to the Gaussian field and is based on a constrained refined Delaunay triangulation, combined with linear basis functions, allowing for more flexibility and resulting in less information loss.

For the sake of the exposition, we here describe the construction of marked point process models based on the simple gridding approach combined with an intrinsic Gauss Markov random field (Rue and Held 2005). The same approach to constructing these or related models with one or more shared random fields may be used based on the SPDE approach. Specifically, for the gridding approach consider $N = n_{row} \times n_{col}$ grid cells $\{s_{ij}\}$, each with area $|s_{ij}|$, $i = 1, \ldots, n_{row}$, $j = 1, \ldots, n_{col}$. The points in a point pattern can then be described by $\{\xi_{ijk_{ij}}\}$ with $k_{ij} = 1, \ldots, y_{ij}$, where y_{ij} denotes the observed number of points in grid cell s_{ij}.

Conditioning on the point pattern and conditionally on $\eta_{ij} = Z(s_{ij})$, we assume y_{ij} follows a Poisson distribution,

$$y_{ij}|\eta_{ij} \sim Po(|s_{ij}|\exp(\eta_{ij})),$$

see Rue et al. (2009), Illian et al. (2012b), and model η_{ij} as

$$\eta_{ij} = \beta_{01} + \sum_{p \in \mathcal{I}} \beta_{1p} z_p(s_{ij}) + \beta_{s1} f_s(s_{ij}) + u_{ij}, \qquad (1)$$

where the function $f_s(s_{ij})$ is a spatially structured effect modelled here as a an intrinsic Gauss Markov random field (CAR) model and reflecting spatial structure in the pattern that cannot be explained by the covariates $\sum_{p \in \mathcal{I}} \beta_{1p} z_p(s_{ij})$ and u_{ij} is a spatially unstructured (i.i.d.) error field. For reasons of identifiability, β_{s1} is typically chosen to be equal to 1.

The marks **m** are assumed to depend on the pattern and are hence sharing the same spatially structured effect $f_s(s_{ij})$. They follow some probability distribution with parameters $\theta_1, \ldots, \theta_q$, i.e. $m(\xi_{ijk_{ij}})|\kappa_{ijk_{ij}} \sim F_\theta(\kappa_{ijk_{ij}}, \theta_1, \ldots, \theta_q)$ with

$$\kappa_{ijk_{ij}} = \beta_{02} + \sum_{p \in \mathcal{I}} \beta_{2p} z_p(s_{ij}) + \beta_{s2} \cdot f_s(s_{ij}) + v_{ijk_{ij}},$$

where $v_{ijk_{ij}}$ is another error term. The two error terms u_{ij} and $v_{ijk_{ij}}$ account for additional structure in the data that cannot be captured by either the covariates or the spatially structured error field that is unique to the spatial pattern and the marks, respectively.

This construction may be generalised to more than one mark (Illian et al. 2012b) and to replicated patterns (Illian et al. 2012a). Illian et al. (2012b) use it to model the dependence between the chemical properties of eucalyptus leaves and the frequency of koalas visiting those trees, while accounting for the spatial structure formed by the trees and hence the autocorrelation in the chemical properties and their proximity. The results indicate that koalas prefer to visit trees whose leaves have a high degree of palatability. Illian et al. (2012a) use a similar approach to distinguish between long-term versus short-term suitability of the local environment for plant species in a sand dune. The health status of plants (mark) showed different spatial structuring and dependencies than the pattern formed by the species, indicating different survival strategies at different time scales.

While these models have clear ecological interpretation and hence relevance, there has been little uptake of the methodology within the ecological community so far. Interacting with potential users through designated courses discussing the methodology and software, as well as the types of models that can be fitted with them, has made it clear to us that potential users can see their relevance. However, it has also highlighted that the practicality and usability of the models has not been discussed enough. The following sections reflect on reasons for this limited uptake. Section 5 focuses on the lack of discussion of the role of the spatially structured field in the literature and the resulting difficulties for potential users in terms of interpretation and prior choice, while Sect. 6 reflects on the lack of complexity in point process models, in particular in terms of complex data structures and observation processes common in ecology.

5 Practicalities of model fitting—the role of the spatial field

Considering the practicality of model fitting implies going beyond the development of relevant models and the provision of software to fit these, through thinking about how the models are used in context to understanding and addressing potential hurdles in practice. In the context of the point process models we have considered here, i.e. log Gaussian Cox processes in a Bayesian context and fitted with the R-INLA library, a clear hurdle for the user is the interpretation of the GMRF that distinguishes them from the models they have come across otherwise. However, random fields are not a concept that most users are familiar with. This is not surprising as it is a mathematical structure that is not necessarily intuitively easy to understand for those largely unfamiliar with spatial statistics, even for trained statisticians. But, what is more, it is not the mathematical concept or the formal definition of a GMRF as such that is directly relevant to the users when they fit a log Gaussian Cox process to their data. Rather, it is the role of the GMRF in the specific models that they have to understand— and that statisticians need to convey in order to provide them with an understanding of the methodology as a whole. Users need to be clear as to why a spatial field is included in a model that otherwise looks very much like the generalised linear models that are commonplace in the ecological literature. This is particularly relevant since many users mainly encounter the models through fitting them in a software package rather than reading about them in a statistical paper or textbook. Effectively commu-
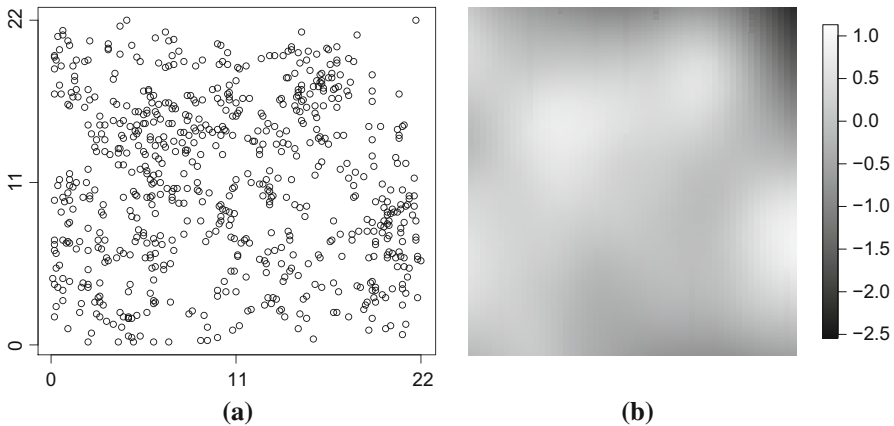
Fig. 1 The spatial pattern formed by individuals of the species *Andersonia heterophylla* (**a**) and the estimated spatial effect from the model in Eq. 2 (**b**) in a 22 m × 22 m observation window

nicating the role of the spatial field to the users implies an explicit discussion among statisticians—a discussion that has been lacking within the statistical literature as well as in the ecological literature.

The traditional role of the spatial field in a point process has been to explain unobserved environmental covariates, often in cases where no covariates were available (Møller and Waagepetersen 2007) and a random field is used to account for spatial structures caused by these unobserved covariates. However, in practice a simple model, such as one where the latent field is of the following form

$$\eta_{ij} = \beta_0 + f_s(s_{ij}) + u_{ij}, \tag{2}$$

with notation as above, provides limited information to the user. The only outcome that fitting this model achieves is an estimated spatial field that can then be plotted. In many cases, this does not provide much more information than a simple plot of the pattern itself. For instance, consider a spatial pattern derived from an Australian plant community (Armstrong 1991). The original data set comprises the locations of 67 different species collected in a 22 m × 22 m square. Here, we consider the pattern formed by one of these plant species, *Andersonia heterophylla* (Fig. 1a).

The estimated spatially structured field in Fig. 1b illustrates how the intensity of the pattern varies through space. However, the information this conveys to the user is mainly a figure that looks very much like a smoothed version of the original point pattern and contains little additional information beyond this. From an ecological perspective, it would be far more interesting to find out why there are more individuals in some part of the plot than in others, i.e. to relate the pattern to spatially varying covariates such as data on soil nutrients to identify those nutrients that are relevant to a particular species (John et al. 2007). For example, in Fig. 1 no data on covariates are available, so we cannot go much further here. However, in many other data sets covariates are available and we might consider a model such as the following, where notation is as in Eq. (1):

$$\eta_{ij} = \beta_0 + \sum_{p \in \mathcal{I}} \beta_p z_p(s_{ij}) + f_s(s_{ij}) + u_{ij}. \tag{3}$$

With this kind of model, we have to rethink the role of the spatial field—thinking about it from the point of view of the users—who would be familiar with a modelling approach that assesses the significance of a number of covariates to the structure of a spatial pattern. From their point of view, the field is an additional term that takes on the role of accounting for spatial structure *unexplained* by existing covariates and it is important to point out what effect this additional term has on inference. At first sight, this new role seems rather similar to that in the model in Eq. (2), as it accounts for spatial structure in a point pattern. However, it does this jointly with the set of covariates, and hence, the new role requires us to reflect on the spatial field in the context of the inference made on the effect of covariate and the interpretation of the modelling results. The properties of the estimated spatial field, in particular its smoothness and its impact on inference, need to be discussed. If the field is too "wiggly", it explains most—and too much—of the spatial structure in the pattern and hence outcompetes any of the covariates, as it can be made arbitrarily "wiggly", which results in overfitting. Similarly, if the spatial field is too smooth, the estimated variance of the coefficients β_p relating to the covariates is underestimated, resulting in spurious significance.

In general, the issue of spatial confounding of a spatially structured random field with covariate coefficients for example in generalised linear mixed models has been discussed (Reich et al. 2006; Hodges and Reich 2010). A potential remedy, restricted spatial regression, forces the spatial field to be orthogonal to the space spanned by the covariates (Reich et al. 2006; Hodges and Reich 2010; Hanks et al. 2015). While this seems appealing, the coefficients of the covariates are given as much variability as possible, in turn leading to spurious significance (Hanks et al. 2015). What is more, the resulting spatial field, and in particular its role suddenly becomes much less clear. Its role is now to account for remaining spatial structure that is *orthogonal* to, i.e. independent of, the existing covariates. In practice, it is very unclear what this remaining spatial structure represents in a biological sense.

There has been very little discussion on how the smoothness of the spatial field impacts inference in the context of point processes. In other words, a question that relates to one of the simplest Cox process models relevant in practice has not been addressed by the statistical literature. A user who aims to fit this kind of model is given little advice as what to do in practice and the effect of model assumptions on the statistical analysis. Illian et al. (2012a) suggest visually comparing the spatial scales at which covariates operate and aiming for a degree of smoothness for the spatial field that does not pick up spatial structures at a smaller scale than that of the covariates. However, this is far from being a thorough or principled approach.

So—what do we need to do to improve this situation? When considering log Gaussian Cox processes in a Bayesian context as we do here, the smoothness of the spatial field is reflected in the hyperpriors that are chosen for the random field—and hence discussing the impact of the smoothness of the spatial field translates into a discussion on appropriate prior choice. There has recently been much discussion on appropriate prior choice for complex Bayesian models, especially in the context of providing default priors as implemented in software (Simpson et al. 2017). This clearly under-

lines the relevance of this discussion in the context of usability as users have to be appropriately advised on the impact of the choice of prior on the statistical analysis and its results. If software is provided along with default priors, a clear discussion of this impact has to initially happen within the statistical community.

In the specific case of spatial point processes and Cox process models, this concerns the prior specifications for the spatially structured field, as the field, as mentioned above, might easily cover up significant covariate effects due to collinearity with the covariates or overfitting. As a result of interaction with potential users in the context of training workshops, a discussion of the issue has begun in the context of models similar to the one in Eq. (3), see, for example, Sørbye and Rue (2014) and Sørbye and Illian (2017).

However, this is not where the discussion of the role of the spatial field in point process models ends. In what we have discussed so far, it is assumed to account for spatial structure due to unavailable covariates. There is no reason why spatial structure in a spatial pattern should be the result of "missing" covariates in all contexts, or for the spatial field to have that role. For example, in the joint models discussed in Sect. 4 the role of the field can be more complex. In the koala model in Illian et al. (2012b), for instance, the spatial field accounts for soil covariates that influence the spatial pattern of trees, but also indirectly both the chemistry of the leaves and the koala frequency—and it also reflects dependence among marks and spatial pattern.

The spatial field does not always have to act as a nuisance term that is included in a model; it might rather be the main interest of a particular study. For example, this might be the case when it is well known that the available set of covariates influences spatial structure and when one aims to identify further potential causes for apparent clustering in a point pattern. The plot of the spatial field may then be used to elicit knowledge in discussions with experts. Similarly, the parameters of the GMRF themselves may be of interest. For instance, assessing differences in the range of a Matérn field in different habitats might reflect biological processes that differ among habitat types; for example, if dispersal limitation varies among habitats (Bakka et al. 2015).

Clearly, in the discussion with users we need to distinguish first- and second-order behaviour explicitly and make it clear that without replicated patterns—which in practice are rarely available—this cannot be done. But we also need to not just mention the spatial field in passing but make it explicit why it is included in the specific model and what it reflects—for instance whether it represents a first order or a second-order process. In all these cases, the varying roles of the spatial field imply that appropriate prior choice for its parameters has to be discussed in context. In particular, it is important to discuss thoroughly the role of the spatial field for a specific model with those that are using the methodology.

6 Future perspectives in the light of more complex models

To finish our considerations, we now provide perspectives on future developments in spatial point process modelling in the light of the focus of this paper, i.e. issues of relevance and practicability related to complex data structures relevant in ecology. Specifically, we discuss this in the context of point pattern data on large spatial scales

and of spatio-temporal point processes. These scenarios are of particular relevance in ecology since large scale spatial and spatio-temporal data sets are becoming increasingly available, and it is important to discuss suitable analysis approaches, see e.g. Gerrodette and Forcada (2005).

6.1 Incomplete detection

As mentioned in Sect. 2, traditionally point process methodology has been developed for either "Window sampling", where the observed point pattern is seen as a small portion of a much larger, infinite pattern that has the same properties as the pattern in the observation window, or for the "Small world Model", which assumes that the observed pattern is a small finite pattern that has been observed in its entirety (Baddeley et al. 2015). Both of these scenarios imply that complete detection of each individual or at most that the pattern has been thinned with a fixed probability p that does not vary in space (referred to as p-thinning in Illian et al. (2008)), an assumption that is often unrealistic in ecology. While, for example, the rainforest data sets discussed in Brown et al. (2013) are relatively large in comparison with those traditionally considered in the point process literature, the relevant observation window is still small enough that it is feasible to collect information on every single tree within the 50 ha plots.

In animal studies, however, a much larger spatial scale is often relevant and surveying the entire area of interest and hence detecting every individual then becomes infeasible. Often data are collected along transects that cover only a small proportion of the area and individuals are less likely to be detected the further away they are from the observer. In other cases, in the context of habitat selection modelling, data on animals are frequently collected by telemetry methods, where only a small proportion of the population has been observed and observations on the same individuals have been made sequentially through time. In these cases, the observation process is dictated by practical data collection issues and needs to be taken into account during modelling. Clearly, to make point process methodology relevant here, models need to include a consideration of the observation, but for spatial point processes, complex observation processes have rarely been discussed. In this context, recent work by Johnson et al. (2010, 2013) has examined point process methods for distance sampling and for telemetry data, respectively.

Yuan et al. (2015) consider data collected by distance sampling along k line transects and use a thinned log Gaussian Cox process to model these. Here, the intensity of potentially observable individuals $\lambda(s)$ is combined with a detection function $g_k(s)$, resulting in a thinned point process with intensity function $\Lambda(s) = \lambda(s)g_k(s)$, ignoring details on temporal aspects of the approach in Yuan et al. (2015) here for simplicity. Specifying and fitting the model has been facilitated by the computational efficiency of INLA, combined with the flexibility of the SPDE approach as introduced by Lindgren et al. (2011), and the emerging software package, `inlabru`, which has been developed with the aim of making the approach in Yuan et al. (2015) and extensions of it accessible to users. As the classical (un-thinned) log Gaussian Cox process can be regarded as a special case of the models discussed in that paper, `inlabru` may

also be used to fit point process models with a simpler observation process; the same is true for other spatial models, in particular geo-referenced models.

In terms of practicality, it is again important to consider the role of the random field in a specific application. Since the spatial pattern is observed at a larger scale, it might represent processes at this large scale that the covariates cannot capture and these can be biotic and abiotic processes. For instance, these processes might refer to larger-scale migration patterns rather than local dispersal limitation. For telemetry data, the spatial field might have a temporal component, capturing the temporal dependence of consecutive observations from the same animal.

6.2 Spatio-temporal models

This brings us to the next and final topic of this paper. The discussion so far has focused entirely on spatial point process modelling. Clearly, in practice, the ecology of individuals also concerns their behaviour in time. In the point process literature, there has been relatively little discussion on spatio-temporal point processes in general. Developments so far have focused mainly on very specific data structures, typically in epidemiology (Diggle et al. 2013), and there have been individual cases of point processes being used in other contexts such as housing construction (Duan et al. 2009) or armed conflict (Zammit-Mangion et al. 2012). This might be partly due to a lack of data and hence little explicit motivation for the construction of models. What is more, considering spatio-temporal models implies an increase in model complexity, leading to increased computational complexity, which is likely to have hindered development. Certainly, however, another complication has been that it is difficult to say what a standard simple spatio-temporal data structure looks like, making it particularly difficult to consider a general class of standard spatio-temporal point process models. Specifically, this concerns, for instance, the issue of object persistence and the relative roles of spatial and temporal structures. The objects represented by the points in a point pattern vary in terms of their persistence relative to the time-scale of observations. For example, individual trees in a forest live for many years and hence many of the trees are observed repeatedly in repeated censuses, while instantaneous events such as earthquakes have little or no duration. Very different issues of dependence among the points need to be considered in these cases. In fact, the spatial dependence of the events might themselves be the subject of interest, for example, whether mortality is spatially structured, as in Bagchi et al. (2011).

A related type of dependence that is likely to exist is that between the spatial and temporal structures in the pattern. While one question here concerns the validity of separability assumptions, the roles of the spatial and temporal structures vary among different studies and so do the relevant types of data aggregation. If there is an interest in changes in the overall spatial structure over time, aggregating the data into patterns in discrete time might be appropriate (Altieri et al. 2015). In contrast, if the global spatial pattern across, say, a number of years is of interest rather than its change through time temporal dependence is merely a nuisance, e.g. the local clustering in space at a small temporal resolution such as in earthquakes. When developing relevant models it is again important to capture the different dependence structures appropriately to

relate the model structure to the relevant scientific question. In terms of practicality, the role of the now spatio-temporal field clearly reflects these different dependence structures.

7 Discussion—the benefits of the dialogue

Some might argue that the role of applied statistics is to provide users with data analysis tools that help them to answer their scientific questions. However, this rather one-directional information flow becomes bidirectional and more interesting once the statistician is seen as an interdisciplinary researcher rather than as a mere method provider. Specifically, in a functioning interdisciplinary collaboration the interaction with the applied subject will contribute to the scientific dialogue on both sides. It will not only provide answers to specific questions from the user community, but also motivate new research questions in both scientific fields. It is obvious that this is the case in statistical research when new data structures resulting, e.g. from emerging technologies require new statistical approaches (c.f. telemetry data). However, an intense interdisciplinary collaboration will also challenge current practice in the context of familiar data structures. It allows the statistical researcher to focus on statistical development in truly relevant areas rather than in areas that seem relevant or interesting from the (naive) statistician's point of view. In addition, the bidirectional dialogue will also challenge common practice in the applied science and help improve the applied users' understanding of the differences and similarities of related methods that they commonly apply, such as generalised additive models (Wood 2006) or those that have made their way into the methodological literature in the field (Golding and Purse 2016).

This paper provides examples of this discussion where the relevance and practicality of the methods have been increased as a result of issues highlighted through interdisciplinary dialogue. We discuss a change in paradigm within a debate on ecological theories of biodiversity, the development of complex marked point process models that answer concrete ecological questions. We also highlight the importance of considering the practical side of model fitting, concerning prior choice and model interpretation, with a focus on rethinking the modelling process, in particular in terms of inference and interpretation of the spatial field.

Clearly, we have so far only discussed a few issues that have prevented a more widespread application of spatial point processes in ecology. For instance, there has been very little discussion in the literature on model comparison and model evaluation for spatial point processes. Model fit is commonly assessed on the basis of a summary characteristic (Møller and Waagepetersen 2003, 2007), which requires repeated simulation from the fitted model. With increasingly complex models, this can become time-consuming and hence impractical. Furthermore, summary characteristics can always only capture one aspect of the spatial structure and for complex models individual summary characteristics cannot capture much. In a more formal approach to model assessment, Baddeley et al. (2005) construct residuals for spatial point process models and Nightingale et al. (2015) consider a reversible jump approach to model choice. Both these approaches are currently restricted to relatively

simple Gibbs processes of unmarked and multi-type point patterns, respectively. However, when fitting other statistical models in practice, ecologists routinely apply model assessment and comparison methodology to assess the quality of their model fit and compare many different models. The statistical literature does not currently provide appropriate methodology in the context of spatial point process models. Even for relatively simple models further developments are needed, beyond the issues discussed above to make point process methods more usable, and eventually more common place as an analysis tool for ecological data.

In summary, it can be useful to see method development and the model fitting process from the point of view of users rather than from the point of view of the mathematical structure. The latter can be useful when developing approaches and when identifying structural similarities to allow approaches to be generalised, but when looking at the necessities of users it might be more useful to initially think in terms of the data problem and the statistical approaches the users are familiar with, or have been using so far, so that methodology is placed in context. For instance, it might be more constructive to think about how spatial point process models are related to generalised linear models, say, and what this implies in practice rather than discussing them as models based on complicated random variables. Similarly, it helps to point out what questions can be answered with the more complex methodology that existing methods cannot answer or can answer less well, rather than pointing out mere mathematical differences. This will require us to think more about the Cox processes' main "selling point", the spatially structured random field, in terms of what it means for the practical interpretation of the models and, inadvertently, its role in a specific application. Now that we are in a situation where methods can be fitted more easily than in the past there is still a lot of work to be done on making the methods usable—they are complex models which implies complex interpretation and complex communication across disciplinary boundaries.

References

Altieri, L., Scott, E.M., Cocchi, D., Illian, J.B.: A changepoint analysis of spatio-temporal point processes. Spat. Stat. (2015). doi:10.1016/j.spasta.2015.05.005

Anderson-Teixeira, K.J., Davies, S.J., Bennett, A.C., Gonzalez-Akre, E.B., Muller-Landau, H.C., Wright, S.J., Abu Salimand, K., Almeyda Zambrano, A.M., Alonso, A., Baltzer, J.L., et al.: CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change. Glob. Change Biol. **21**(2), 528–549 (2015)

Armstrong, P.: Species patterning in the heath vegetation of the Northern Sandplain. Honours thesis, University of Western Australia (1991)

Baddeley, A., Turner, R.: Practical maximum pseudolikelihood for spatial point processes. N. Z. J. Stat. **42**, 283–322 (2000)

Baddeley, A., Turner, R., Møller, J., Hazelton, M.: Residual analysis for spatial point processes (with discussion). J. R. Stat. Soc. Ser. B **67**, 617–666 (2005)

Baddeley, A., Rubak, E., Turner, R.: Spatial Point Patterns: Methodology and Applications With R. CRC Press, Boca Raton (2015)

Baddeley, A.J., Turner, R.: Spatstat: an R package for analyzing spatial point patterns. J. Stat. Softw. **12**, 1–42 (2005)

Bagchi, R., Henrys, P.A., Brown, P.E., Burslem, D.F.R.P., Diggle, P.J., Gunatilleke, C.V.S., Gunatilleke, I.A.U.N., Kassim, A.R., Law, R., Noor, S.: Spatial patterns reveal negative density dependence and habitat associations in tropical trees. Ecology **92**(9), 1723–1729 (2011)

Bakka H., Illian J.B., Simpson D., Rue H.: Log-Gaussian Cox processes with spatially varying second order properties. In: Proceedings of the 30th International Workshop on Statistical Modelling, Linz 2015. (2015)

Baldeck, C.A., Harms, K.E., Yavitt, J.B., John, R., Turner, B.L., Valencia, R., Navarrete, H., Davies, S.J., Chuyong, G.B., Kenfack, D., et al.: Soil resources and topography shape local tree community structure in tropical forests. Proc. R. Soc. Lond. B Biol. Sci. **280**(1753), 20122532 (2013)

Brooks, T.M., Mittermeier, R.A., da Fonseca, G.A.B., Gerlach, J., Hoffmann, M., Lamoreux, J.F., Mittermeier, C.G., Pilgrim, J.D., Rodrigues, A.S.L.: Global biodiversity conservation priorities. Science **313**(5783), 58–61 (2006)

Brown, C., Law, R., Illian, J.B., Burslem, D.F.R.P.: Linking ecological processes with spatial and non-spatial patterns in plant communities. J. Ecol. **99**(6), 1402–1414 (2011)

Brown, C., Burslem, D.F.R. P., Illian, J.B., Bao, L., Brockelman, W., Cao, M., Chang, L.W., Dattaraja, H.S., Davies, S., Gunatilleke, C.V.S.: Multispecies coexistence of trees in tropical forests: spatial signals of topographic niche differentiation increase with environmental heterogeneity. Proceedings of the Royal Society of London B: Biological Sciences, 280 (1764):20130502 (2013)

Brown, C., Illian, J.B., Burslem, D.F.R.P.: Success of spatial statistics in determining underlying process in simulated plant communities. J. Ecol. (2015). doi:10.1111/1365-2745.12493

Chape, S., Harrison, J., Spalding, M., Lysenko, I.: Measuring the extent and effectiveness of protected areas as an indicator for meeting global biodiversity targets. Philos. Trans. R. Soc. B Biol. Sci. **360**(1454), 443–455 (2005)

Chave, J.: Neutral theory and community ecology. Ecol. Lett. **7**(3), 241–253 (2004)

Condit, R., Pitman, N., Leigh Jr., E.G., Chave, J., Terborgh, J., Foster, R.B., Nunez, P., Aguilar, S., Valencia, R., Villa, G., Müller-Landau, H.C., Losos, E., Hubbell, S.P.: Beta diversity in tropical forest trees. Science **295**, 666–669 (2002)

Connell, J.H.: On the role of natural enemies in preventing competitive exclusion in some marine animals and in rainforest trees. In: den Boer, P.J., Gradwell, G.R. (eds.) Dynamics of Populations, pp. 298–313. Centre for Agricultural Publishing and Documentation, Wageningen (1971)

Cornulier, T., Bretagnolle, V.: Assessing the influence of environmental heterogeneity on bird spacing patterns: a case study with two raptors. Ecography **29**(2), 240–250 (2006)

Debski, I., Burslem, D.F.R.P., Lamb, D.: Ecological processes maintaining differential tree species distributions in an australian subtropical rain forest: implications for models of species coexistence. J. Trop. Ecol. **16**(03), 387–415 (2000)

Debski, I., Burslem, D.F.R.P., Palmiotto, P.A., Lafrankie, J.V., Lee, H.S., Manokaran, N.: Habitat preferences of Aporosa in two Malaysian forests: implications for abundance and coexistence. Ecology **83**(7), 2005–2018 (2002)

Decan, A., Mens, T., Claes, M., Grosjean, P.: On the development and distribution of R packages: an empirical analysis of the R ecosystem. In: Proceedings of the 2015 European Conference on Software Architecture Workshops (2015)

DeSoto, L., Olano, J.M., Rozas, V., De la Cruz, M.: Release of Juniperus thurifera woodlands from herbivore-mediated arrested succession in Spain. Appl. Veg. Sci. **13**(1), 15–25 (2010)

Diggle, P.J.: Statistical Analysis of Spatial Point Patterns, 2nd edn. Hodder Arnold, London (2003)

Diggle, P.J., Moraga, P., Rowlinson, B., Taylor, B.M.: Spatial and spatio-temporal log-Gaussian Cox processes: extending the geostatistical paradigm. Stat. Sci. **28**(4), 542–563 (2013)

Duan, J.A., Gelfand, A.E., Sirmans, C.F.: Modeling space-time data using stochastic differential equations. Bayesian Anal. **4**(4), 733–758 (2009)

Ferraro, P.J., Pattanayak, S.K.: Money for nothing? A call for empirical evaluation of biodiversity conservation investments. PLoS Biol. **4**(4), 482 (2006)

Gerrodette, T., Forcada, J.: Non-recovery of two spotted and spinner dolphin populations in the eastern tropical pacific ocean. Mar. Ecol. Progress Ser. **291**, 1–21 (2005)

Geyer, C.J., Møller, J.: Simulation procedures and likelihood inference for spatial point processes. Scand. J. Stat. **21**, 359–373 (1994)

Golding, N., Purse, B.V.: Fast and flexible Bayesian species distribution modelling using Gaussian processes. Methods Ecol. Evol. **7**, 598–608 (2016)

Goreaud, F., Pélissier, R.: Avoiding misinterpretation of biotic interactions with the intertype K_{12}-function: population independence vs. random labelling hypotheses. J. Veg. Sci. **14**(5), 681–692 (2003)

Gunatilleke, C.V.S., Gunatillekeand, I.A.U.N., Esufali, S., Harms, K.E., Ashton, P.M.S., Burslem, D.F.R.P., Ashton, P.S.: Species-habitat associations in a Sri Lankan dipterocarp forest. J. Trop. Ecol. **22**(04), 371–384 (2006)

Haase, P.: Spatial pattern analysis in ecology based on Ripley's K-function. J. Veg. Sci. **6**, 575–582 (1995)

Hamill, D.N., Wright, S.J.: Testing the dispersion of juveniles relative to adults: a new analytic method. Ecology **67**, 952–957 (1986)

Hanks, E.M., Schliep, E.M., Hooten, M.B., Hoeting, J.A.: Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. Environmetrics (2015). doi:10.1002/env.2331

Harms, K.E., Condit, R., Hubbell, S.P., Foster, R.B.: Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. J. Ecol. **89**(6), 947–959 (2001)

Harper, J.L.: Population Biology of Plants. Academic press, New York, London, San Francisco (1977)

Hodges, J.S., Reich, B.J.: Adding spatially-correlated errors can mess up the fixed effect you love. Am. Stat. **64**, 325–334 (2010)

Högmander, H., Särkkä, A.: Multitype spatial point patterns with hierarchical interactions. Biometrics **55**(4), 1051–1058 (1999)

Hubbell, S.P.: Tree dispersion, abundance and diversity in a tropical dry forest. Science **203**, 1299–1309 (1979)

Hubbell, S.P.: The Unified Neutral Theory of Biodiversity and Biogeography. Monographs in Population Biology, vol. 32. Princeton University Press, Princeton (2001)

Illian, J.B., Burslem, D.F.R.P.: Contributions of spatial point process modelling to biodiversity theory. Journal de la Société Française de Statistique **148**, 9–29 (2007)

Illian, J.B., Benson, E., Crawford, J., Staines, H.J.: Multivariate methods for spatial point processes—a simulation study. In: Baddeley, A., Gregori, P., Mateu, J., Stoica, R., Stoyan, D. (eds.) Spatial Point Process Modelling and Its Applications, pp. 125–130. Publicacions de la Universitat Jaume I, Castelló de la Plana (2004)

Illian, J.B., Penttinen, A., Stoyan, H., Stoyan, D.: Statistical Analysis and Modelling of Spatial Point Patterns. Wiley, Chichester (2008)

Illian J.B., Sørbye S.H., Rue H., Hendrichsen D.: Using INLA to fit a complex point process model with temporally varying effects–a case study. J. Environ. Stat. **3**, 1–25 (2012a)

Illian, J.B., Sørbye, S.H., Rue, H.: A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). Ann. Appl. Stat. **6**(4), 1499–1530 (2012b)

Inlabru Development Team. The inlabru software (2017). www.inlabru.org

Janzen, D.H.: Herbivores and the number of tree species in tropical forests. Am. Nat. **104**, 501–528 (1970)

John, R.C., Dalling, J.W., Harms, K.E., Yavitt, J.B., Stallard, R.F., Mirabello, M., Hubbell, S.P., Valencia, R., Navarrete, H., Vallejo, M., Foster, R.B.: Soil nutrients influence spatial distributions of tropical tree species. Proc. Natl. Acad. Sci. USA **104**, 864–869 (2007)

Johnson, D.S., Laake, J.L., Ver Hoef, J.M.: A model-based approach for making ecological inference from distance sampling data. Biometrics **66**(1), 310–318 (2010)

Johnson, D.S., Hooten, M.B., Kuhn, C.E.: Estimating animal resource selection from telemetry data using point process models. J. Anim. Ecol. **82**(6), 1155–1164 (2013)

Jones, J.P.G., Collen, G., Atkinson, P.W.J., Baxter, P., Bubb, J., Illian, B., Katzner, T.E., Keane, A., Loh, J., McDonald-Madden, E.: The why, what, and how of global biodiversity indicators beyond the 2010 target. Conserv. Biol. **25**(3), 450–457 (2011)

Law R., Dieckmann U.: Moment approximations of individual-based models. In: The geometry of ecological interactions: simplifying spatial complexity, pp. 252–270 (2000)

Law, R., Dieckmann, U., Metz, J.A.J.: Introduction. In: Dieckmann, U., Law, R., Metz, J.A.J. (eds.) The Geometry of Ecological Interactions: Simplifying Spatial Complexity, pp. 1–6. Cambridge University Press, Cambridge (2000)

Law, R., Illian, J.B., Burslem, D.F.R.P., Gratzer, G., Gunatilleke, C.V.S., Gunatilleke, I.A.U.N.: Ecological information from spatial patterns of plants: insights from point process theory. J. Ecol. **97**, 616–628 (2009)

Ledo, A., Illian, J.B., Schnitzer, S.A., Wright, S.J., Dalling, J.W., Burslem, D.F.R.P.: Lianas and soil nutrients predict fine-scale distribution of above-ground biomass in a tropical moist forest. J. Ecol. **104**(6), 1819–1828 (2016)

Lindgren, F., Rue, H.: Bayesian spatial and spatiotemporal modelling with r-inla. J. Stat. Soft. **63**(19), 1–25 (2015). doi:10.18637/jss.v063.i19

Lindgren, F., Rue, H., Lindström, J.: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. **73**(4), 423–498 (2011)

Losos, E.C., Leigh, E.G.: Tropical Forest Diversity and Dynamism: Findings From a Large-Scale Plot Network. University of Chicago Press, Chicago (2004)

Magurran, A.E., McGill, B.J. (eds.): Biological Diversity: Frontiers in Measurement and Assessment. Chapman & Hall, London (2010)

Magurran, A.E.: Ecological Diversity and Its Measurement. University Press, Cambridge (1988)

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J.L., He, F.: Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecol. Lett. **10**(10), 995–1015 (2007)

Møller, J., Waagepetersen, R.P.: An introduction to simulation-based inference for spatial point processes. In: Møller, J. (ed.) Lecture Notes in Statistics, vol. 137, pp. 143–198. Springer, New York (2003)

Møller, J., Waagepetersen, R.P.: Statistical Inference and Simulation for Spatial Point Processes. Chapman & Hall/CRC, Boca Raton (2004)

Møller, J., Waagepetersen, R.P.: Modern statistics for spatial point processes (with discussion). Scand. J. Stat. **34**, 643–711 (2007)

Møller, J., Syversveen, A.R., Waagepetersen, R.P.: Log Gaussian Cox processes. Scand. J. Statist. **25**, 451–482 (1998)

Murrell, D.J., Law, R.: Heteromyopia and the spatial coexistence of similar competitors. Ecol. Lett. **6**(1), 48–59 (2003)

Myllymäki, M., Penttinen, A.: Conditionally heteroscedastic intensity-dependent marking of log Gaussian Cox processes. Stat. Neerl. **63**, 450–473 (2009)

Myllymäki, M., Grabarnik, P., Seijo, H., Stoyan, D.: Deviation test construction and power comparison for marked spatial point patterns. Spat. Stat. **11**, 19–34 (2015)

Naeem, S., Bunker, D.E., Hector, A., Loreau, M., Perrings, C.: Biodiversity, Ecosystem Functioning, and Human Wellbeing. An Ecological and Economic Perspective. Oxford University Press, Oxford (2009)

Nightingale, G.F., Illian, J.B., King R.: Pairwise interaction point processes for modelling bivariate spatial point patterns in the presence of interaction uncertainty. J. Environ. Stat. **7**(3) (2015)

Ogata, Y., Tanemura, M.: Estimation of interaction potentials of marked point-patterns through the maximum-likelihood method. Biometrics **390**, 315–338 (1985)

Pélissier, R., Goreaud, F.: Ads package for R: a fast unbiased implementation of the K-function family for studying spatial point patterns in irregular-shaped sampling windows. J. Stat. Softw. **63**(6), 1–18 (2015)

Pocheville A.: The ecological niche: history and recent controversies. In: Heams, T. et al. (eds.) Handbook of Evolutionary Thinking in the Sciences, pp. 547–586. Springer, Netherlands (2015)

Purves, D.W., Law, R.: Heteromyopia and the spatial coexistence of similar competitors. Ecol. Lett. **6**, 48–59 (2003)

Rands, M.R.W., Adams, W.M., Bennun, L., Butchart, S.H.M., Clements, A., Coomes, D., Entwistle, A., Hodge, I., Kapos, V., Scharlemann, J.P.W.: Biodiversity conservation: challenges beyond 2010. Science **329**(5997), 1298–1303 (2010)

Reich, B.J., Hodges, J.S., Zadnik, V.: Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. Biometrics **62**, 1197–1206 (2006)

Rue, H., Held, L.: Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall, London (2005)

Rue, H., Martino, S., Chopin, N.: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). J. R. Stat. Soc. B **71**, 319–392 (2009)

Schlather, M.: On the second-order characteristics of marked point patterns. Bernoulli **7**, 99–117 (2001)

Schmid, B., Balvanera, P., Cardinale, B.K., Godbold, J., Pfisterer, A.B., Raffaelli, D., Solan, M., Srivastava, D.S.: Biodiversity, Ecosystem Functioning and Human Wellbeing: An Ecological and Economic Perspective. Oxford University Press, Oxford (2009)

Seidler, T.G., Plotkin, J.B.: Seed dispersal and spatial pattern in tropical trees. PLoS Biol **4**(11), e344 (2006)

Simpson, D.J., Illian, J.B., Lindgren, F., Sørbye, S.H., Rue, H.: Going off grid: computationally efficient inference for log-gaussian cox processes. Biometrika **103**(1), 49–70 (2016)

Simpson, D.P., Rue, H., Riebler, A., Martins, T.G., Sørbye, S.H.: Penalising model component complexity: A principled, practical approach to constructing priors. Stat. Sci. **32**(1), 1–28 (2017)

Slik, J.W.F., Arroyo-Rodríguez, V., Aiba, S.-I., Alvarez-Loayza, P., Alves, L.F., Ashton, P., Balvanera, P., Bastian, M.L., Bellingham, P.J., van den Berg, E., et al.: An estimate of the number of tropical tree species. Proc. Natl. Acad. Sci. **112**(24), 7472–7477 (2015)

Sørbye, S.H., Rue, H.: Scaling intrinsic Gaussian Markov random field priors in spatial modelling. Spat. Stat. **8**, 39–51 (2014)

Sørbye, S.H., Illian, J.B., Simpson, D.P., Burslem, D.F.R.P.: Careful prior specification avoids incautious inference for log-Gaussian Cox point processes. (in submission) (2017)

Studeny, A.C., Buckland, S.T., Illian, J.B., Johnston, A., Magurran, A.E.: Parametric evenness indices: a goodness-of-fit approach. 2, (2010). doi:10.1890/ES10-00074.1

Tippmann, S.: Programming tools: adventures with R. Nature **517**(7532), 109–110 (2015)

Tittensor, D.P., Walpole, M., Hill, S.L.L., Boyce, D.G., Britten, G.L., Burgess, N.D., Butchart, S.H.M., Leadley, P.W., Regan, E.C., Alkemade, R.: A mid-term analysis of progress toward international biodiversity targets. Science **346**(6206), 241–244 (2014)

van Lieshout, M.: Markov Point Processes and Their Applications. Imperial College Press, London (2000)

Watt, A.S.: Pattern and process in the plant community. J. Ecol. **35**, 1–22 (1947)

Wiegand, T., Moloney, K.T. (eds.): Handbook of Spatial Point-Pattern Analysis in Ecology. Chapman & Hall, Boca Raton (2013)

Wiegand, T., Moloney, K.A.: Rings, circles, and null-models for point pattern analysis in ecology. Oikos **104**, 209–229 (2004)

Wiegand, T., Gunatilleke, S., Gunatilleke, N., Okuda, T.: Analyzing the spatial structure of a Sri Lankan tree species with multiple scales of clustering. Ecology **88**(12), 3088–3102 (2007)

Wood, S.: Generalized Additive Models: An Introduction With R. CRC Press, Boca Raton (2006)

Wright, J.S.: Plant diversity in tropical forests: a review of mechanisms of species coexistence. Oecologia **130**(1), 1–14 (2002)

Yuan, Y., Bachl, F.E., Lindgren, F., Borchers, D.L., Illian, J.B., Buckland, S.T., Rue, H., Gerrodette, T.: Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales (in submission) (2015)

Zammit-Mangion, A., Dewar, M., Kadirkamanathan, V., Sanguinetti, G.: Point process modelling of the Afghan War Diary. Proc. Natl. Acad. Sci. **109**(31), 12414–12419 (2012)

Zillio, T., Condit, R.: The impact of neutrality, niche differentiation and species input on diversity and abundance distributions. Oikos **116**(6), 931–940 (2007)

Zuur, A.F., Ieno, E.N., Elphick, C.S.: A protocol for data exploration to avoid common statistical problems. Methods Ecol. Evol. **1**(1), 3–14 (2010)