doi: 10.1111/j.1467-9469.2007.00571.x
 Vol 34: 685–711, 2007

Discussion of 'Modern Statistics for Spatial Point Processes'

ABSTRACT. The paper 'Modern statistics for spatial point processes' by Jesper Møller and Rasmus P. Waagepetersen is based on a special invited lecture given by the authors at the 21st Nordic Conference on Mathematical Statistics, held at Rebild, Denmark, in June 2006. At the conference, Antti Penttinen and Eva B. Vedel Jensen were invited to discuss the paper. We here present the comments from the two invited discussants and from a number of other scholars, as well as the authors' responses to these comments. Below Figure 1, Figure 2, etc., refer to figures in the paper under discussion, while Figure A, Figure B, etc., refer to figures in the current discussion. All numbered sections and formulas refer to the paper.

The Editors

Comment by Antti Penttinen (University of Jyväskylä)

Modern point process statistics allows the use of covariates for controlling spatial variation in (marked) point patters, uses the likelihood inference and steps towards Bayesian modelling have been taken. The authors have had many initiatives in this computationally intensive science. They are in the forefront of putting forward new developments in point process statistics that follow the current mainstream of statistical thinking. This is a long way forward from the works by Svedberg (1922) and Matérn (1960, 1971), the two earliest Scandinavian contributors of spatial point process statistics.

My first comment is towards the role of marks. In this work a marked point process $\mathbf{\Phi}=\{u, m_u : u \in \mathbf{X}\}$ has been considered as an unmarked point process in a higher-dimensional space as was done in the modelling of the Norwegian spruce data. This kind of thinking is feasible in mathematics but not in application-oriented statistics. In general, the marks m_u have a role of characterizing the points, but they are not coordinates. One consequence is, for example, that when defining stationarity the marks are preserved in the translations of locations.

Let us continue with the Norwegian spruce modelling. As mentioned in the paper the regular feature in the unmarked point pattern is due to man-made thinning of the forest according to forester's rule such as: 'The wanted stem number is 500–600 per hectare and the trees are evenly distributed'. This rule is just for avoiding competition! Further, the size of a tree may affect the retaining probability of a tree in the thinning; hence, size-dependent marks can be affected as well. Hence, the interpretation of the dependence of marks is not obvious. The concept 'influential zone', introduced by the authors, is welcomed. It gives a refinement for the standard mark correlation analysis (Stoyan & Stoyan, 1995). A simple new device based on influental zone type reasoning could be

$$A(k) = \mathbf{E}\left(\sum_{i \neq j} |b(u_i, km_i) \cap b(u_j, km_j)|\right).$$

Here, m_j is the diameter at breast height of the tree at u_j, and k is a prefixed 'dilation factor' ($k=5$ in the paper). An application of this summary supports my earlier view, which was based on mark-correlation analysis, that there is probably no correlation between the marks, or if any, it is very weak (cf. Fig. A).

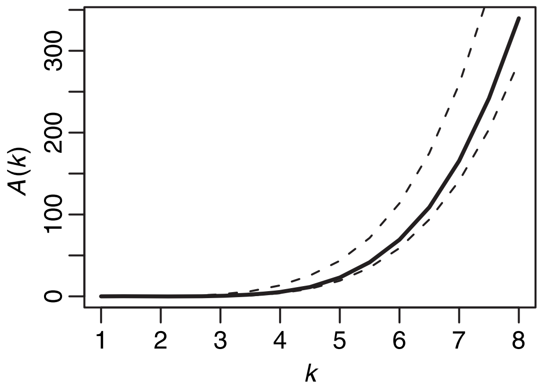

Fig. A. The estimate of $A(k)$ (solid curve) and pointwise maximum and minimum envelopes (dotted curves) from 99 simulations under independent marking, conditional on the point locations. Zero boundary has been applied.

My third remark is on the 'well-defined' inhomogeneous K- and g-functions. These are essentially defined for the class of second-order reweighted stationary processes as are the two inhomogeneous models, log Gaussian Cox process and generalized Thomas process, presented in the paper. But what would be the devices of examining this property in data? Using locally defined pair-correlation functions, I found empirical evidence that this property does not hold for the tropical rainforest data.

My fourth critical comment is on the new approach of defining residuals for a point process model fit through a nice application of the conditional intensity and the Ruelle–Nguyen–Zessin formula. This is without doubt a deep theoretical contribution. Before celebrating this result, I am waiting applications where this new residual analysis really gives advantages over the more traditional summary-based model critique.

Something very traditional can be found in the paper: The window problem is ignored as is the case in most applications of point process statistics. The window may be crucial both for modelling and interpretation of the results, for example, in plant biology and physics. In some applications it may have a similar role as site-dependent covariates.

Finally I would add some personal views over the prospects for spatial point process statistics. We are still lacking a practice of analysing complex marked point pattern data with covariates and generic models for marked point patterns. Maybe the development in point process statistics has been too conventional? One could possibly find means such as statistical reconstruction of point patterns through summary-constrained optimization (Tscheschel & Stoyan, 2006). Also new types of data would benefit the development. An example is the use of additional hereditary information when deciding the mother–offspring relationship in cluster models as in Shimatani (2002). Third, the traditional models, often originated from physics, are models in equilibrium while this is seldom true for biological systems. Incorporating temporal aspects (see, e.g. Särkkä & Renshaw, 2006), or even spatio-temporal data, would lead to modelling with improved interpretation.

Comment by Eva B. Vedel Jensen, Michaela Prokešová and Gunnar Hellmund (University of Aarhus)

First of all, we want to congratulate the authors for writing a very clear review of the state of the art of modelling and analysing spatial point process data. Next, we will like to comment on some additional issues concerning spatio-temporal modelling, inhomogeneity, Cox processes and simulation-free procedures.

The examples considered in the paper are mainly used for illustrative purposes. We agree with the authors that 'many scientific problems call for new spatial point process methodology for analysing complex and large data sets (often with marks and possibly in space-time)'. To the list of such application areas mentioned in the paper we can add neuroscience where each point of a *marked spatio-temporal point process* is used to model the time of onset, duration and spatial extent of a neuronal activation at a particular spatial position. New challenges are here to construct spatio-temporal clustered point processes with long-distance dependencies and to develop efficient Bayesian analysis of huge data sets with spatio-temporal point process models as priors (Taskinen, 2001; Hartvig, 2002; Jensen & Thorarinsdottir, 2007). It could have been worthwhile to use a few more pages on future aspects, especially relating to spatio-temporal point processes.

Specific procedures are discussed for introducing *inhomogeneity* into the different types of point process models considered. In the case of Cox processes, log-linear inhomogeneity is specified multiplicatively in the random intensity function or, equivalently, as a location-dependent thinning of a homogeneous Cox process. There is a variety of other ways of introducing inhomogeneity depending on the type of Cox process considered. In the case of a shot-noise Cox process, one obvious alternative possibility is to let the process Φ be inhomogeneous. We are happy that collaboration has been started between the authors and us concerning the statistical analysis of such an inhomogeneous Cox process. In Fig. B, different possibilities for introducing inhomogeneity into a shot-noise Cox process are illustrated by simulation. In the case of a log Gaussian Cox process, inhomogeneity may be obtained by a transformation of a stationary Gaussian process (Monestiez *et al.*, 1993). For the Gibbs point processes, the focus is on first-order inhomogeneity (non-constant first-order potential). Alternative ways of introducing inhomogeneity in a Gibbs point process are briefly discussed in section 10. In Jensen & Nielsen (2001), a review of four such classes of inhomogeneous

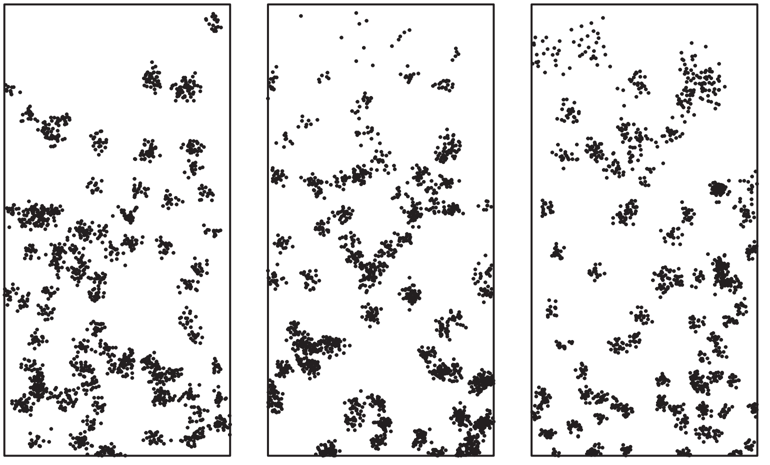

Fig. B. Examples of realizations of inhomogeneous shot noise Cox processes. The (mean) intensity function is the same in all three examples. To the left, the inhomogeneity is introduced via the location of the mothers in Φ, in the middle the inhomogeneity is obtained by a location dependent thinning of a homogeneous Cox process while to the right the inhomogeneity is obtained by a local scaling mechanism.

Gibbs point processes are given, together with a solution of the non-trivial problem of introducing log-linear inhomogeneity into these models.

Two main classes of *Cox processes* are discussed separately, viz. the shot-noise Cox processes and the log Gaussian Cox processes. We think that it is very much worthwhile to introduce a unified framework, thus being able to see shot-noise and log Gaussian Cox processes in a new light, investigate their relationships and define further natural extensions, including models with both shot-noise and log Gaussian components. The latter type of extension would, in particular, be of interest for the tropical rain forest data. Such unified framework can be based on the concept of a Lévy basis (infinitely divisible and independently scattered random measure); cf. Barndorff-Nielsen & Schmiegel (2004) and references therein. The basic ideas are studied in depth in Hellmund (2005), spatio-temporal point processes are discussed in Prokešová *et al.* (2006b), and an extended version (Hellmund *et al.*, 2007) is in preparation. A Lévy driven Cox process (LCP) has driving field of the form

$$\Lambda(u) = \int_{\mathcal{R}} k(r, u) L(dr), \quad u \in S,$$

where L is a positive Lévy basis on $(\mathcal{R}, \mathcal{B})$, $\mathcal{R}$ is a bounded subset of $\mathbb{R}^d$, say, and k is a non-negative weight function on $\mathcal{R} \times S$. The Lévy–Ito representation implies that the resulting process is a shot-noise Cox process with trend. A log Lévy driven Cox process (LLCP) has driving field of the form

$$\Lambda(u) = \exp\left(\int_{\mathcal{R}} k(u, r) L(dr) \right).$$

It turns out that Λ is a product of a log Gaussian field and an independent log shot noise field. The extension to spatio-temporal point processes is straightforward; cf. Prokešová *et al.* (2006b). Lévy bases also appear to be a powerful tool in modelling of turbulence and growth phenomena (Barndorff-Nielsen & Schmiegel, 2004; Jensen *et al.*, 2007; Jónsdóttir *et al.*, 2007).

We suggest that the *simulation-free estimation procedures* are studied further. As mentioned by the authors, the estimating function (47) is the one obtained in a likelihood-based approach under a Poisson model with $\rho(u)$ as intensity function. In the paper, a next step is to estimate the remaining model parameters by using an estimating function based on a second-order characteristic like the K-function. As an alternative, a two-step likelihood procedure may be considered in analogy with the ones used for statistical inference in transformation and local scaling point process models; cf. Nielsen & Jensen (2004) and Prokešová *et al.* (2006a). To be more specific, let the model parameters be denoted by $\theta = (\alpha, \omega)$, where α parametrizes the intensity function ρ_α, say, and ω is the remaining part of the parameter. Write the likelihood as

$$L(\theta) = L_1(\alpha) L_2(\theta),$$

where L_1 is the likelihood function for a Poisson point process with intensity function ρ_α and L_2 is the remaining part of the likelihood function. The idea is now to calculate an estimate $\hat{\alpha}$ of α and subsequently estimate ω by maximizing $L_2(\hat{\alpha}, \omega)$ with respect to ω. Note that L_2 is the likelihood function of the data with respect to the inhomogeneous Poisson point process with intensity function ρ_α.

Comment by Adrian Baddeley (University of Western Australia and CSIRO)

The authors give an excellent summary of the state of the art in statistical methods for spatial point pattern data. It is an exciting time to be working in this field. Two decades ago, it was widely believed that spatial point patterns are not amenable to standard methods of statistical inference. It is now clear that the fundamental tools of statistics, such as likelihoods, residuals and inferential techniques, are abundantly applicable and effective for spatial point pattern data.

As the scope of applications expands, we encounter larger and more complex data sets, more complex statistical issues and practical problems.

Modelling of spatial inhomogeneity (section 10.2) is an important topic requiring further development. The authors' work on inference for Cox processes (sections 7.3 and 7.4) is a key contribution. I hope (12) will become known as the 'Waagepetersen model'; it seems very important. A challenging problem is to deal with 'singular' inhomogeneity, such as the concentration of points almost exactly along a line, observed with earthquakes and galaxy surveys (Baddeley, 2006; Ogata & Katsura, 1988).

Computationally intensive, simulation-based inference will play an increasing role (section 10.4). One might expect that statistically inefficient estimating equations and pseudolikelihood inference will eventually disappear from favour. However, the same advances in computer technology that enable statisticians to perform simulation-based inference also enable scientists to collect enormous data sets. We will probably always need a plurality of techniques, ranging from the statistically inefficient but computationally fast ones, up to the statistically optimal but computationally expensive.

Techniques described in this paper concern the analysis of a single, spatial pattern of points. Data sets consisting of many spatial point patterns arise, for example, in microscopy, where each microscope field of view generates an observed point pattern. Statistical theory for such data is still under development (Diggle *et al.*, 2000; Bell & Grunwald, 2004; Baddeley *et al.*, 2007).

Statisticians usually assume a spatial point pattern has been observed without error. In reality the observations may be subject to sampling bias and measurement error. It is scientifically crucial to take into account all sources of sampling bias. The observed spatial distribution of astronomical gamma-ray bursts must be corrected for the uneven sky coverage of the observing satellite and its localization errors. Having personally observed Minke whales in the wild, I wonder whether their natural curiosity may lead to positive sampling bias (over-representation relative to the true abundance) and positive correlation between Minke sightings.

The interpretation of summary statistics requires more sophistication. Consider the classical interpretation of the pair correlation function as indicating either randomness, attraction or repulsion. Like any second-moment property, the pair correlation does not completely justify such causal interpretations. Defensible inference about interpoint interaction requires a model of interaction. This is one argument for retaining Gibbs/Markov models. I wonder whether the difficulties of fitting scale parameters in current Markov models (section 10.3) are attributable to the non-differentiability of the interaction terms.

Much interesting work remains to be done in spatial statistics. The new residuals for spatial point patterns (section 6) should be extended to partial residuals, leverage diagnostics and similar tools. Space–time processes are another exciting frontier.

Comment by Jean-Michel Billiot (Université Grenoble II)

It is true that spatial statistics recently have made significant advances. Its use in a large field of areas calls for more and more flexible statistical models to describe the largest kind of situations. Of course, several models such as the Gibbs models come from statistical physics; refer, among others, the Lennard–Jones interaction but also the Widom–Rowlinson model

(Widom & Rowlinson, 1970) for the liquid vapour phase transition. As it is pointed out in the paper, in the attractive case, things are quite unsatisfactory and often more difficult. As an example, consider the Ruelle class of superstable and regular potential (Ruelle, 1970). The attractive part needs to be balanced by a repulsive one at a small scale so as to satisfy the stability assumption. Notice that, for this model, some characteristics are known such as a fine bound of the correlation functions, the Kirkwood Salsburg equations take a simple form, and ergodic theorems, cluster expansion, variational principle and large deviations inequalities are available (Nguyen & Zessin, 1979; Georgii, 1994). That explains partly why some fine asymptotic results on maximum-likelihood estimation (MLE) or maximum-pseudolikelihood estimation (MPLE) can be established for this family (Mase, 1992, 1999). Unfortunately, it becomes harder for general attractive Gibbs models. For these models, the local stability assumption useful for geometric convergence of spatial birth and death process and also for perfect simulation is not satisfied. But Ruelle (1970) shows how it can be weakened to obtain the thermodynamic limit. Clearly, for models with a finite range, we can avoid edge effect with a minus sampling, but even in that case the stability of the energy ensuring the finiteness of the partition function remains a problem, for example, in the case of Quermass or nearest neighbours interactions (Kendall *et al.*, 1999; Bertin *et al.*, 1999). It is easy to understand that the stationarity assumption is not reasonable in various applications and it is a challenge to develop theory to make better statistical analyses. One way could be trying powerful large deviations techniques and maybe derive asymptotic properties of estimators for some kind of inhomogeneous models. Otherwise, determinantal and permanent point processes are promising – they come from quantum mechanics whose explanatory power is matchless. But it is a non-local theory, so finding Gibbsian properties of such models seem to be a hard question. However, the knowledge of the Papangelou conditional intensity may be helpful to do statistical inference at least for the pseudolikelihood method.

Comment by Noel Cressie (The Ohio State University)

The authors have written a very clear, comprehensive paper that I enjoyed reading. I would like to pursue one aspect mentioned in their review, that of *aggregation* of spatial point processes.

Spatial lattice processes for at most a countable number of 'small areas' are often assumed to be Markov random fields (MRFs). This is a very attractive way to build in spatial dependence based on the collection of all univariate conditional distributions of the ith small area's value (Y_i, say) conditional on all other values $\{Y_j : j \neq i\}$. The 'spatial Markov' feature of these models comes from positing a neighbourhood N_i of small areas whose values determine the ith conditional distribution (e.g., Besag, 1974) referred to above. Specifically, if $[U|V]$ denotes the conditional probability density (or mass function) of U given V, then an MRF satisfies

$$\left[Y_i \mid \{Y_j : j \neq i\}\right] = \left[Y_i \mid \{Y_j : j \in N_i\}\right],$$

for all small areas i.

Consider the bounded region S and let $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ denote the point pattern of $n \geq 0$ locations in S. For example, these might be home addresses of people who died of colon cancer in a given time period. Confidentiality considerations mean that the data are released in an aggregated form, as counts. Let $S = \cup A_i$, where $\{A_i\}$ denote the small areas over which colon cancer deaths are aggregated. Define

$$Y_i \equiv \#\left(\cup\{x_j : x_j \in A_i\}\right),$$

where $\#(B)$ denotes the number of elements in the finite set B. These types of count data $\{Y_i\}$, and MRF models for them, are common in the disease-mapping literature (e.g., Banerjee *et al.*, 2004, section. 5.4). In fact, to make meaningful epidemiological statements, one requires further demographic data $\{E_i\}$, the number of people at risk by small area. This requirement is immaterial to my question, which I shall now ask.

Are there spatial point process models for $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ that result in common MRF models for $\{Y_i\}$ or $\{Y_i/E_i\}$? An affirmative response to this question would be important, because epidemiological relationships modelled at the individual level could be inferred from data and resulting models at the small-area level. Curiously, there are a few results in the other direction, giving spatial point processes that result from infill asymptotics of MRFs (Besag *et al.*, 1982; Cressie *et al.*, 2000). But this is far less interesting than the connection from point processes to lattice models under aggregation.

Any further light the authors can shine on *aggregation* of spatial point processes would generate a lot of interest in both applications and theory.

Comment by Pavel Grabarnik (Russian Academy of Sciences) and Aila Särkkä (Chalmers University of Technology)

The paper is a comprehensive review of spatial point process models as well as methods and tools for model checking, simulation and inference. The examples studied are all from biology, and they include features like clustering, inhomogeneity and interaction.

When modelling (biological) point pattern data, several aspects such as biological reasoning, model convenience and aims of modelling should be taken into account. For example, when modelling clustering, one should think what is the reason for clustering. As mentioned in the paper, it can be, for instance, seed dispersal or variation in intensity of some covariate. These two lead to different choice of a model. However, it may be more desirable to choose, for example, a likelihood-based model to make the inference easier than a biologically most relevant model. Furthermore, sometimes the goal is to generate point patterns with specific statistics, and then the biological relevance of the generating mechanism (model) is less important.

If the main emphasis is on finding a realistic model, biological reasoning cannot be ignored. For example, if the interaction we are modelling is known to be non-symmetric, we should fit a model capable of describing a non-symmetric character of the interactions. It seems natural in example 5.2 that the locations of Messor nests affect the locations of Cataglyphis nests but not vice versa. Interaction between neighbouring trees in a forest is also non-symmetric or hierarchical: A small tree does not affect a large tree as much (if at all) as the large tree affects the small one. However, trees that are approximately of the same size affect each other. We suggest dividing the trees into size classes and apply a hierarchical model (Grabarnik & Särkkä, 2007). Trees in the class of largest trees are affected only by other trees in the same class, but trees in smaller size classes, on the other hand, are affected by other trees in the same class and by all larger trees. Interaction within classes is symmetric, and interaction between the classes hierarchical. Both a traditional Gibbs point process model and our model with hierarchical interactions may be able to produce patterns with similar statistics and patterns that are similar to the data. However, interpretation of the between-classes interaction is different in the two cases: In the former case the interaction function describes the mutual interaction between the trees, while in the hierarchical case it describes the effect of large trees on small ones. Therefore, the hierarchical model is more relevant from the point of view of ecology.

Comment by Yongtao Guan (Yale University)

I would like to thank the authors for writing up such a comprehensive summary paper. This paper with no doubt will serve as an excellent reference for graduate students who wish to work on topics related to spatial point processes in their dissertation research. My main comments are about modelling inhomogeneous spatial point patterns, which are detailed as follows.

For the *Beilschmiedia pendula Lauraceae* data studied in this paper, the authors fit two different models: a log Gaussian Cox process (LGCP) model and an inhomogeneous Thomas process (ITP) model. Both models appear to fit the data well, as can be observed by looking at Fig. 9. However, the inferential results regarding the key regression parameters are very different. Therefore, an important question is which model is more appropriate for the given data.

It should be noted that the LGCP and the ITP models have very different ecological implications. Specifically, the LGCP model suggests that the inhomogeneity is caused solely by the spatially varying environmental conditions (both observed and unobserved), whereas the ITS model attributes it to both the (observed) environmental conditions and (unobserved) dispersal effects. In reality, the inhomogeneity may be caused by a combination of many factors such as spatially varying environmental conditions (both observed and unobserved), dispersals and interactions among trees. It is generally difficult to disentangle effects from these factors (which is what the LGCP and the ITS models attempt to do) as they cannot be quantified by observed data. Guan & Loh (2007) recently proposed a thinned block bootstrap procedure that can be used to make inference on the regression parameters without assuming a specific parametric model (e.g. LGCP or ITP) on the high-order structures of the process. The main assumption is that the process is second-order reweighted stationary. There is a great need for such flexible non-parametric inferential procedures because of the complex structures that many spatial point pattern data often possess.

My last comment is regarding how to assess the goodness-of-fit of a fitted intensity model for clustered data. I strongly agree with the authors that Gibbs process and its variations are not appropriate for such data. For model diagnostics, Waagepetersen (2005) suggested defining the residuals in terms of the intensity function. However, we also need to account for the variability of the obtained residuals. This is necessary as the variance of the residuals can be highly variable, regardless how well the regression parameters can be estimated. More research is needed on this topic.

Comment by Peter Guttorp (University of Washington)

I very much enjoyed reading this paper, and congratulate the authors for having produced a very nice overview of places where point process work has gone over the last decade or two.

Many of the examples in this paper deal with ecological problems. Ecology really deals with interactions, often a multitude of interactions at different scales. Traditionally, statistical ecology has been cursed by using the Poisson process as a null hypothesis. While the Poisson adequately describes the idea that ‘nothing is going on’, rejecting it does not help understanding ecological systems.

Data nowadays are collected on many scales and in many ways, for example, in the case of forest ecology using quadrant counts (relatively quick), satellite data of different coverage and resolution (arriving continuously, at least on the time scale of slow ecological change) and detailed mapping (slow and expensive). A hierarchic structure, using the true map as an underlying unobserved state, is a natural approach. But the intensity is often scale-dependent, having, for example, components of clustering (from seed dispersion), repulsion (from crowns blocking sunlight) and non-stationarity (from fertility gradients and orography). Would a reasonable approach be to overlay these different processes in the style of a generalized additive model (GAM)?

One often wants to establish statistical (or causal) relationships between measurement systems, to use cheaper and less labour-intensive measurement methods to draw conclusions regarding changes in spatial patterns. While this can be thought of as a space–time problem, it is really more of a comparison of patterns to detect change. How would the authors go about testing for change of this type?

Can one use the whale analysis to develop improved survey designs?

The method of moments and other estimating equations have been very common in point process models in hydrology. Of course, it is not efficient, and can be rather disastrously bad. In the rain forest example, the estimating equation values of σ and β_3 are outside the corresponding Bayes intervals. Is this due to high variability in the estimating equation estimators or due to dependence on the prior?

In hydrology, where I have done most of my point process work, various types of cluster processes are very common, as they are somewhat natural when modelling precipitation events (although the underlying process of storm fronts rarely would be Poisson, as fronts by definition need to be separated). It would be useful to have a complete list of the cluster processes for which Markov chain Monte Carlo (MCMC) methods (or exact likelihood methods) have been worked out.

I would like to finish with two historical remarks. First, what is the origin of the Poisson process? I have not found anything in Poisson's work that relates to it. However, a paper of Clausius (1858) in developing the kinetic theory of heat uses the assumptions of a Poisson process to calculate the mean free path of a particle in three dimensions. So far this is the earliest calculation I have found of this type. The second remark refers to the Neyman–Scott process. While their 1958 cosmology paper is a common original reference, the clustering process originates in Neymans 1939 paper on contagious distribution, preceding Thomas' special case of Gaussian dispersion function by over a decade.

Comment by Olle Häggström (Chalmers University of Technology)

My congratulations to Jesper Møller and Rasmus Waagepetersen for their well-written and authoritative account of an area that has grown explosively during the last decade or two. I have two comments on their paper.

1. Markov chain Monte Carlo (MCMC) convergence. In section 9.2, Møller and Waagepetersen discuss Metropolis–Hastings algorithms for point processes defined in terms of conditional intensities, stressing that local stability of the point process guarantees that the algorithm converges to equilibrium geometrically fast, which in turn implies a central limit theorem (CLT) for Monte Carlo errors. I entirely agree with this statement, but wish to stress that applying it requires a good deal of caution, the reason being that, even if the convergence is geometrically fast, the mixing time may nevertheless be very large.

A prototype example is single-site dynamics for the Ising model in a large two-dimensional box in the phase coexistence region of the parameter space. Such a Markov chain is irreducible and aperiodic with finite state space, and so geometric convergence to equilibrium is automatic. Yet, the time taken to come close to equilibrium becomes astronomical even for moderately sized systems. To make things worse, the system quickly settles down into a kind of metastable behaviour where all the usual MCMC diagnostics instruments will fool us into believing that convergence to equilibrium and CLT behaviour have kicked in. See, for example, Marcelli & Martinelli (1996) for more on this type of metastability phenomena in the Ising model.

The Widom–Rowlinson and area-interaction point processes, studied, for example, in Chayes *et al.* (1995) and Häggström *et al.* (1999), seem bound to exhibit qualitatively the same kind of behaviour in certain parts of the parameter space, although rigorously they are less well-understood. In general, it can be quite difficult to determine whether a particular model leads to such metastable behaviour.

In my view, it is not enough to rely on geometric ergodicity when applying Metropolis–Hastings chains and other MCMC algorithms (including simulated tempering). We also need explicit quantitative control on the mixing time.

2. Modelling. What is the purpose of statistical analysis of spatial point patterns? Sometimes, perhaps, one is content with estimating, say, the intensity function or the pair correlation function, and that is the end of the story. Much more interesting is the situation where the statistical analysis is part of the project of trying to understand, on a deeper level, the physical or biological mechanism underlying the observed point pattern. In such a situation, I feel that the following two general modelling strategies are worth employing.

First, the specific physics or biology of the situation at hand should play a central role in the development of the stochastic point process models. Whatever we know about the underlying physics or biology that appears to influence what kind of point patterns arise should be taken into account in the modelling. This mindset may be contrasted against one where real-world data serve mainly as decorations of the models and methods under consideration. Of course, both attitudes can be found within the research area surveyed by Møller and Waagepetersen, but their selection of material suggests that the latter is still more common.

Second, understanding the phenomena underlying the point pattern typically involves understanding the dynamics by which it is produced. Hence, it is highly desirable that the spatial point process model is a projection of a spatio-temporal model that describes this dynamics. The comment in section 10.3 of the paper advising against Gibbs point process modelling in forestry, on the grounds that growth of forests does not exhibit the time reversibility that the dynamical models producing Gibbs point processes do, is very much in this spirit.

Comment by Ute Hahn (Universität Augsburg)

It is a great pleasure to congratulate Jesper Møller and Rasmus Waagepetersen on their brilliantly written, most interesting and comprehensive overview on recent developments in point process statistics.

As stated in section 10.2, one of the branches currently coming to the fore is the analysis of inhomogeneous point patterns. With a thorough analysis of a tropical rain forest map including concomitant information on the terrain, the authors convincingly illustrate the power of Bayesian analysis for inhomogeneous Cox process models.

While Cox point processes are indeed very versatile models for patterns that exhibit clustering or no interaction, neither these nor second-order intensity-reweighted stationary processes appear to be appropriate for a situation shown in Fig. C of a pattern that is highly regular or repulsive both in the areas with high and low intensity.

Obviously, such patterns cannot even approximately be explained by independent thinning, and therefore, statistical methods based on the second-order intensity-reweighted stationarity assumption are not appropriate in this case.

Section 10.2 mentions other models that have been suggested for inhomogeneous point processes, among others, transformation of homogeneous Markov point processes or local scaling. These models are more appropriate for patterns as the one depicted in Fig. C. I should like to add that, similar to the suggestions made in section 6.2, these approaches also evoke recommendations for their statistical analysis. Nielsen (2000) fits a transformed Strauss process to an inhomogeneous pattern of cells in stomach tissue after having estimated

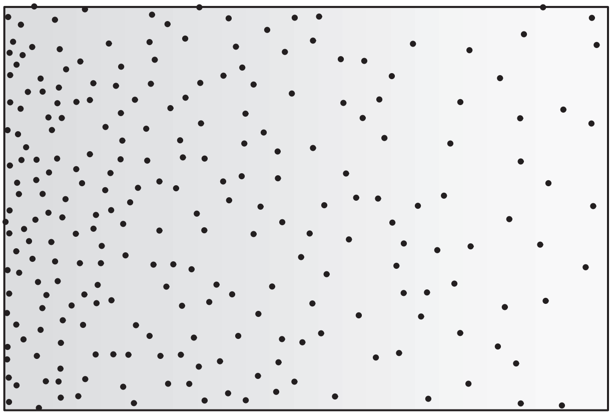

Fig. C. Realization of a locally scaled hard core model.

the backtransformation, and Fleischer *et al.* (2006) apply a similar principle to the clustered inhomogeneous pattern of root profiles in a soil transect and thus obtain second-order statistics including the J-function.

Prokešová *et al.* (2006a) propose an inhomogeneous K-function for point processes with a local scaling property. It is indeed possible to derive locally scaled analogues to all distance based summary statistics, including the F-, G- and J-function, just by replacing the distance $||u-v||$ between points that occurs in the definition of the statistics, for example (41) for the K-function, with the intensity adjusted distance $d_\rho(u,v) := \int_{[u,v]} \sqrt{\rho(s)}ds$, where $[u,v]$ denotes the line segment spanned by u and v. This means that the point process is locally rescaled to unit rate, and further occurrences of the intensity function in the definition of the summary function have to be replaced by unity. The calculation of intensity adjusted distances can be simplified, using averages, to

$$\hat{d}_\rho(u,v) = \frac{1}{2}\left(\sqrt{\rho(u)} + \sqrt{\rho(v)}\right)||u-\boldsymbol{v}||$$

or

$$\hat{d}_\rho(u,v) = 2\left(\sqrt{1/\rho(u)} + \sqrt{1/\rho(v)}\right)^{-1}||u-\boldsymbol{v}||.$$

The latter was preferred in the aforementioned paper for model immanent reasons. A locally scaled K-function can then be defined as

$$K_A^*(t) = \frac{1}{\mu(A)}\mathrm{E}\sum_{u\in\mathbf{X}_A}\sum_{v\in\mathbf{X}\setminus\{u\}}\mathbf{1}[d_\rho(u,v)\le t].$$

Ideally, $K_A^*(t)$ should not depend on the choice of $A\subset\mathbb{R}^2$, and in my experiences it is virtually independent of A even if the model is only close to being 'locally scaled'.

Lothar Heinrich (University of Augsburg)

Jesper Møller and Rasmus P. Waagepetersen have presented a very thorough and comprehensive review of the current state of spatial point process statistics. They have made a commendable and wholly successful effort to demonstrate the great variety of applications and sources of point process modelling in different fields with interesting examples of real life point patterns and by referring to the relevant monographs as well as to original papers of point process literature. The authors avoid the use of the measure theoretic machinery and deeper excursions into point process theory and statistical physics which make this survey paper readable even for statisticians with modest mathematical background. However, the authors do not hide the mathematical rigorousness which is necessary for doing a correct statistical treatment of point process models and mention also theoretically difficult issues such as the problem of phase transitions of Gibbsian point processes on unbounded regions. It is quite natural that the different topics touched in this survey paper are not equally weighted. So main emphasis is put on research areas in point process statistics to which the authors have contributed a lot. In particular, the authors' research on point process models specified by a (random) intensity function (in section 4) or by the Papangelou conditional intensity (in section 5) in connection with the development of computational methods including MCMC techniques to their statistical inference have provided important tools for analyzing especially inhomogeneous point patterns on a bounded sampling region.

From the methodoligical view point it is noteworthy that the approaches based on parametrized (conditional) intensity functions allow to benefit from analogies with generalized linear models and random effect models. This shows that the adaptation of well-established satistical methods including time series analysis to the statistics of spatial point processes can be fruitful perhaps also for issues like experimental design, missing data and outlier analysis, etc.

As mentioned in the introductory section, in practice statisticians are often faced with rather small (mostly planar) sampling windows or with small numbers of points forming a single realization of a possibly inhomogeneous point process. Small means here less than 100 points in a squre $[0,10] \times [0,10]$. It seems that the term *modern statistics* suggests that statistics of such point patters is now possible. Nevertheless, even homogeneous and motion-invariant point processes remain rather complicated stochastic models and large sample or, equivalently, large domain statistics is an appropriate tool for their statistical analysis. In particular, goodness-of-fit tests (of the Kolmogorov-Smirnov and Cramér-von Mises type) based e.g. on Ripley's K-function, see Heinrich (1991), or its scaled versions, see Heinrich (2007), or on other non-parametric *summary statistics* should also be an object of future research aiming at rigorous statistical model identification.

In summary, this survey is well done, it will stimulate future research, and thus Jesper Møller and Rasmus P. Waagepetersen deserve our thanks and congratulations.

Comment by Wilfrid S. Kendall (University of Warwick)

I very much enjoyed reading this paper: the authors are to be congratulated on producing an excellent overview of the current state of art for statistics for point processes – a topic to which they are themselves major contributors. Here are a couple of discussion comments.

1. My attention was caught by the remarks in sections 5.3 and 10.3 concerning phase transition behaviour. The unsuitability of the Strauss model for handling attractive interaction ultimately derives from, and finds extreme expression in, the remarks of Kelley & Ripley (1976), who show that the Strauss local specification fails to produce a well-defined point process when the interaction is positive and the number of points is unconstrained. But phase transition phenomena cause problems for modellers using the attractive area-interaction point process. This presents a rather sharper problem; the attractive area-interaction point process itself exists, and indeed is Ruelle-stable; nevertheless the phase transition exhibited by the process defined on all of $\mathbf{R}^2$ produces a bi-modality for bounded-window processes. What diagnostics would indicate when such phase transition phenomena might be a potential problem for a general point process model (possibly conditioned by data)? Both bi-modality and substantial dependence on boundary conditions would cause slow-down in the coupling from the past algorithms of Kendall & Møller (2000), which may be suggestive for useful diagnostics.

2. Also in section 10.3, the authors comment on issues arising when the pattern to be modelled does not arise from a dynamical process in statistical equilibrium. The issue is subtle. Valid Markov point process models for such situations may themselves be achieved as statistical equilibria of spatial birth–death processes; there is a difference between algorithmic time and model time! However, and clearly, the model cannot itself be justified as arising from an equilibrium argument (though one could justify it empirically in terms of forming an exponential family using a relevant sufficient statistic). The point is reminiscent of Kingman (1977)'s note on Poisson distributions and reproducing populations. To conclude on a note of wild speculation, perhaps it might be possible to establish a result that shows that simple dynamic population models (such as might arise in forestry, to take the example in the paper) must give rise to Markov point process models that involve sufficient positive interaction that phase transition issues might be an issue.

Comment by Andrew Lawson (University of South Carolina)

Drs Møller and Waagepeteresen are to be congratulated on a very nice review of the current state of spatial point process analysis.

General Comments: First of all I was concerned that the authors could only demonstrate the use of spatial point process (SPP) methods for ecological examples. There are many examples of SPPs in spatial epidemiology, bioinformatics, engineering, defence studies and social science. Inclusion of a range of applications would help to convince the wary reader of the benefits of SPP analysis. In my experience, most researchers, *within* and *without* the subject of statistics, shy away from point processes as a vehicle for analysis and it would be useful to try to 'sell' the analysis by its wide application. This underselling is also underscored by the lack of accessible and easy-to-use software for SPP analysis. While some exists in R or S currently (e.g. SPLANCS, SPATSTAT), there is no flexible and simple-to-use packaged software for much of the methodology discussed in this review. In addition, often researchers wish to ask more sophisticated questions about their data rather than simple descriptions of pattern (see the Minke Whale comment below). The flexible modelling of linear predictors, hierarchical modeling with (non-spatial) random effects, missingness (except as edge effects) and model goodness-of-fit are a few areas that are not developed here.

Specific Comments:
Berman–Turner integration schemes: I have worked with these over time and found the simple Dirichlet weight schemes to be very inaccurate (Lawson, 1992). Delauney schemes are better but must be augmented by subtriangles usually to a depth of four to five subdivisions (Wang & Lawson, 2006).

Residuals for PPs: While the authors have recently developed residual analysis for a range of processes, the original suggestion of comparison of a saturated intensity estimate to a model estimate was made for Poisson SPPs in Lawson (1993).

Minke Whale example: The authors analyse data from a dynamic mobile population as a static spatial problem. Both observation position and the population of whales are changing location over time, but the authors present a static analysis. In addition the authors do not present convincing arguments for using their methods compared to others (e.g. Neyman–Scott process; Cowling, 1998; or Hedley & Buckland, 2004). How do their shot-noise Cox process add to our understanding of the underlying processes, compared to other (simpler) methods?

Bayesian approaches: There has been extensive use of Bayesian models for clustering of point and line processes in a range of applications that are not mentioned by the authors (see, e.g. Lawson, 2000; Lawson & Denison, 2002; McKeague & Loiseaux, 2002; Gangnon, 2006; Lawson *et al.*, 2007).

Comment by Jorge Mateu (Universitat Jaume I)

I enjoyed reading this stimulating and timely review paper, which is both certainly interesting and useful for the practice of spatial point process modelling. It is thus a pleasure to congratulate the authors for an able and comprehensive research on available statistical tools for point process analysis.

Particularly welcome in this account is the exposition of the various worked-out examples and applications, the conciseness and the non-technical introduction to the modern theory, and the continuous analogies with generalized linear models and random effect models. This will surely attract the interest of many applied researchers coming from a wide range of disciplines.

Consider first the specialized case of marked point processes, when the marks are modelled through a (quite general) harmonic decomposition of the form

$$Y(t) = c + \sum_{w \in \Gamma} a_w \cos(2\pi wt) + b_w \sin(2\pi wt)$$

where $\Gamma \subset \mathbb{R}$ is a set of frequencies, and $c, a_w, b_w \in \mathbb{R}$, $\forall w \in \Gamma$. Nothing is mentioned in the paper in this case, and the spectral approach can provide a useful statistical tool by considering the marked periodogram $I(p)$, and defining a residual measure, called *discrepancy* (Renshaw *et al.*, 2007). The discrepancy function is related to some of the residuals for point processes commented in the paper, and applying statistical inference is of primary interest.

Although many studies of spatial point processes analyse only point patterns in terms of purely spatial relationships, in real life, point configurations may evolve dynamically over time and space. The understanding of the space–time interdependency is clearly crucial if we are to understand the underlying mechanisms that generate such evolving structures. The authors comment in several parts of the paper on the modelling of cluster processes, and just in one occasion on space–time point processes. In this context we would like to reinforce the use of modern statistical tools when analysing such situations, for instance, (multi-)generations or growing and reproducing particles of spatial point processes.

Ecological and biological interactions usually occur at specific locations promoting the redistribution of organisms in space. This can generate complex ecological systems resulting in different ecological spatial patterns. Multi-generation point processes have the potential to generate spatial point patterns evolving through discrete time. These point processes can be a reasonable approximation for population reproducing at discrete intervals of time and where usually generations are non-overlapping. Diggle (2003) formulates a Cox process where the events of the $(t+1)$th (i.e. offspring) generation are determined by the t-th parent distribution. Whilst Comas & Mateu (2007a) extend this general Cox process formulation to a multi-generation process assuming interaction effects between parent-to-parent, parent-to-offspring and offspring-to-offspring.

In addition, it is also of interest the development and statistical analysis of spatially explicit marked point process models to generate spatial patterns of reproducing and moving cells evolving through continuous time. In this context, marked points (i.e. cells) can divide and move as a result of (a) their own growth and division motions, and (b) the division motions and the growth of their touching neighbours. This provides a modelling framework to simulate cell aggregate (i.e. tissue) dynamics. Moreover, adding movement to dividing and growing points opens up new areas of application as well as new theoretical problems such as the analysis and generation of dense packing of discs (or spheres); cf. Comas & Mateu (2007b).

These two aforementioned situations are built upon particular classes of cluster point processes. Thus, developing strategic inferential tools for the analysis of space–time dynamics and interdependencies is of crucial present (not future) interest.

Comment by Håvard Rue, Sara Martino (Norwegian University of Science and Technology) and Nicolas Chopin (CREST-LS and ENSAE, Paris)

The authors are to be congratulated for providing a nice, modern overview of statistics for spatial point processes, which would not be what it is today without all the contributions from the authors themselves.

We would like to focus our comment on Bayesian inference for log Gaussian Cox processes (LGCP). In particular, we would like to expand on the authors' comment in section 10.4: '…it may be possible to compute accurate approximations of posterior distributions without MCMC'. It is a pleasure to announce that this is now indeed the case! We find our new results extremely exciting, with respect to accuracy, speed and generality (Rue *et al.*, 2007).

We consider the problem of approximating posterior marginals in latent Gaussian models. These cover a large class of models and are characterized by a posterior of the following form:

$$\pi\left(\boldsymbol{x},\boldsymbol{\theta} \mid \boldsymbol{y}\right) \propto \pi(\boldsymbol{\theta})\pi\left(\boldsymbol{x} \mid \boldsymbol{\theta}\right)\prod_{i\in\mathcal{I}}\pi\left(y_i \mid x_i,\boldsymbol{\theta}\right).$$

Here, some data $\boldsymbol{y}$ observe the latent Gaussian field $\boldsymbol{x}$ pointwise. (This assumption can be relaxed.) The likelihood and the covariance matrix of $\boldsymbol{x}$ are parametrized by a typically low-dimensional vector $\boldsymbol{\theta}$. We approximate the posterior marginals for each x_i as follows: first, construct approximations to $\pi(x_i \mid \boldsymbol{\theta},\boldsymbol{y})$ and $\pi(\boldsymbol{\theta} \mid \boldsymbol{y})$, and then, combine them via numerical integration

$$\tilde{\pi}(x_i \mid \boldsymbol{y}) = \sum_j \tilde{\pi}(\boldsymbol{\theta}_j \mid \boldsymbol{y}) \times \tilde{\pi}(x_i \mid \boldsymbol{\theta}_j,\boldsymbol{y}) \times \Delta_j$$

to provide approximations to $\pi(x_i \mid \boldsymbol{y})$. Here, both $\tilde{\pi}(\boldsymbol{\theta}_j \mid \boldsymbol{y})$ and $\tilde{\pi}(x_i \mid \boldsymbol{\theta}_j,\boldsymbol{y})$ are constructed from the classical Laplace approximation, and Δ_j are integration weights. Laplace approximations are tightly connected with Gaussian approximations to full conditionals, which have been frequently used in block Markov chain Monte Carlo (MCMC) algorithms; see, for example, Rue & Held (2005) for an overview. Computational aspects are very important and not straightforward; see Rue *et al.* (2007) for details.

We have re-analysed the tropical rain forest tree data using a slightly different model. The log intensity $\boldsymbol{\eta}$ takes the following form,

$$\eta_i = \beta_0 + \beta_1 \times \text{altitude}_i + \beta_2 \times \text{gradient}_i + u_i + v_i, \quad i = 1,\ldots,200\times 100$$

where $\boldsymbol{u}$ is a second-order (polynomial) intrinsic Gaussian Markov random field, constructed to mimic a thin-plate spline (Rue & Held, 2005, chapter 3) and $\boldsymbol{v}$ is a vector of independent, zero mean, Gaussians. The hyperparameters $\boldsymbol{\theta}$ are the unknown log precisions for $\boldsymbol{u}$ and $\boldsymbol{v}$. As $\boldsymbol{u}$ is (polynomial) intrinsic, we impose the constraint $\mathbf{1}^T\boldsymbol{u}=0$ to separate out the mean intensity. Note that, in this case, the latent field is $\boldsymbol{x}=(\beta_0,\beta_1,\beta_2,\boldsymbol{u}^T,\boldsymbol{\eta}^T)^T$ with dimension 40,003.

The obtained posterior marginals for β_1 and β_2 are shown in Fig. D, with 95% intervals $[0.009, 0.164]$ and [3.78, 9.53] for β_1 and β_2, respectively. We verified our results with long MCMC runs on a cruder dimension, and were unable to detect any approximation error at all. Computing the marginals for the βs took about 10 minutes on a 2.1 GHz laptop. If the

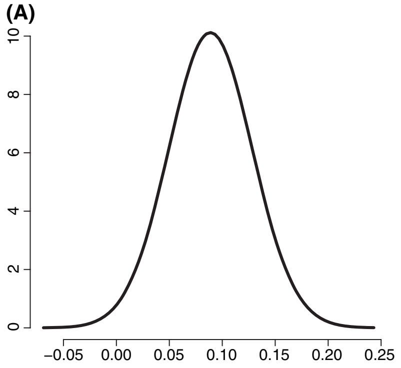

Fig. D. Approximated posterior marginals for β_1 (A) and β_2 (B).

posterior marginals for the spatial component, $\boldsymbol{u}$, are also calculated, the computational cost increases to about 4 hours.

Our results are slightly different from those obtained by the authors. Two reasons may explain this. Firstly, we use a different spatial model, which includes unstructured effects. Secondly, MCMC for such a large model is difficult, not just because of the large dimension of the spatial component itself, but because of the strong correlation between the latent field and its hyperparameters $\boldsymbol{\theta}$. Unless these two components are updated jointly, there is a considerable danger of not exploring the whole space, and hence of underestimating the variability; refer to Rue & Held (2005, chapter 4) for a throughout discussion on this effect.

We end with some final comments on the model chosen for this example. The authors insist on using a continuous spatial model, but note that the exponential covariance function used is not recommended by Stein (1999), because of the implied non-differentiability. Further, the range and variance parameters are both not identifiable under infill-asymptotics (Zhang, 2004), which has implications also for MCMC algorithms. The assumption of linear effects of the covariates should always be verified, for example, by using smooth covariate effects instead; see, for example, Lang & Brezger (2004).

Comment by Frederic P. Schoenberg (University of California, Los Angeles)

I congratulate the authors on an excellent survey of important results surrounding spatial point processes and am delighted to have been asked to contribute to this discussion. I was particularly impressed with the way in which the authors place modern point process models and methods within a context of general statistical methodology so that point process methods might be easily understood by statisticians in general.

One question I have for the authors is: how should one deal with the problem of interactions between the covariates, that is, when the loglinear form in (7) or (8) is not appropriate? For instance, in example 4.1, what if the intensity of tropical rain forest trees at a given location depended also on the interaction between the altitude and gradient at that location? How might one test whether such an interaction is present, and if so, how can this interaction be incorporated into the modelling and estimation procedure?

Although the residual plots considered in the paper are very nice and potentially useful, my sense is that the story is not yet finished when it comes to residual methods and goodness-of-fit testing for spatial point processes. Most of the techniques considered in the paper seem primarily designed to address whether the model's parametric form seems generally adequate, but how should one test whether the process is best modelled as a shot-noise process versus a Gibbs process? The most useful of the residual methods described in the paper is probably the smoothed residual field, but even it seems to be a rather low-power diagnostic, and it is rather difficult to imagine any reasonably flexible model demonstrating gross signs of inadequacy with such a plot. The diagnostic tests in section 6.2 seem powerful for describing second-order properties, but it seems unlikely that any of the residual methods described in section 6.1 could be used to discriminate between a Poisson process and a process with moderate short-range clustering, which is unfortunate.

As someone who deals mostly with spatial–temporal (rather than purely spatial) models for point processes, I could not help but feel that something was lacking in the analyses of some of the example data sets. Obviously these data sets were picked with simplicity in mind, but it nevertheless should be pointed out that to ignore the evolution of points in time in the case of analysing whale sightings or ant nests hardly seems advisable in practice.

In addition to the proposed extensions listed in section 10, another area that may become important is the extension of existing models and techniques to the case of non-simple point processes. If two or more points may be overlapping at exactly the same location, then several of the formulas and interpretations considered in the paper become invalid, and models for such non-simple processes seem to be elusive in the current literature. For plant locations, non-simplicity might not be a problem, but one could imagine two whale sightings at different times occurring at essentially the same location (up to the scale at which measurements are recorded). For the temporal marked point process case, some extensions to non-simple processes were described in Schoenberg (2006). It would be interesting for spatial extensions to be considered as well.

Again, my congratulations to the authors on their extraordinarily clear, thorough and well-written review.

Comment by Dietrich Stoyan (TU Bergakademie Freiberg)

This paper presents a wealth of powerful methods, and I would like to congratulate the authors on their excellent work. However, it will take time for the methods to become widely used in the sciences. In addition to the didactically well-chosen examples of the authors, more instructive examples are needed. And these examples could be analysed more thoroughly, as I explain here for the rain forest data.

Already visual inspection of Fig. 2 shows that it is a clustered pattern, no K-function is necessary to come to this conclusion. The important problem of determining the range of correlation r_{corr} is not discussed, which is perhaps impossible by means of the K-function as shown in Fig. 9. Assuming stationarity I calculated the pair correlation function (p.c.f.) for the bandwidths $h = 0.5, 1.2$ and 5 m. The three estimates practically coincide and suggest a $r_{\text{corr}} \approx 150$ m, which may correspond to a cluster diameter of 300 m. This scale may correspond to local interaction between plants and the environment. By the way, an argument as that of the authors against p.c.f. estimation could favour the empirical distribution function over histograms, as these are 'sensitive to the choice of' the class interval length. To my surprise, the estimates $\hat{\sigma} = 1.33$ and $\hat{\alpha} = 34.7$ given in section 8.2 used for a stationary log Gaussian Cox process yield a very good approximation of the empirical p.c.f.

If stationarity for the pattern is not assumed, both $\hat{K}(r)$ in Fig. 9 and my $\hat{g}(r)$ tell only little about interaction among trees, as these estimates result from averaging over the whole window. Therefore, subwindows should be considered. Of particular interest are 'quasi-homogeneous' subwindows, which resemble research plots of foresters in some way. They give information on local interaction of points, with reduced influence of covariates. I considered the two subwindows $W_1 = [560, 680] \times [0, 120]$ and $W_2 = [580, 760] \times [180, 410]$. The corresponding intensities and (local) p.c.f.s differ greatly. While clustering is indicated for W_1, the pattern in W_2 seems to be close to complete spatial randomness or Poisson process behaviour. The estimated r_{corr} are 20 m and 5 m, respectively, marking the scale of local interaction between single trees resulting from the individual's size and dispersal. The different results for W_1 and W_2 seem to reveal that the assumption of second-order reweighted stationarity (s.o.r.s.) may not hold with the given window or is not better than the stationarity assumption. I miss in the paper any discussion of the validity of the s.o.r.s. assumption.

Finally, I doubt that point process statistics is the right approach for treating the 'question whether the intensity of the trees may by viewed as a spatially varying function of the covariates'. I would recommend a geostatistical approach using a constructed regionalized variable $Z(x)$ with $x \in \mathbb{R}^2$. A possible form is

$$Z(x) = N(b(x, R)),$$

that is, $Z(x)$ is the number of points in the disk of radius R centred at x. I calculated the variogram for $R = 12.5\,\text{m}$ using a $25 \times 25\,\text{m}^2$ lattice of 800 points. It looks like an empirical variogram as often observed in geostatistical studies and suggests $r_{\text{corr}} \approx 500\,\text{m}$. This scale seems to correspond to the variability of the covariates shown in Fig. 3.

Figures of the p.c.f. and of the variogram mentioned can be found in an extended version of this comment at www.mathe.tu-freiberg.de/Stoyan/e-publik.html.

Response by Jesper Møller and Rasmus P. Waagepetersen (Aalborg University)

We thank the discussants for their enthusiastic and thought-provoking remarks.

Data examples: Our data examples were chosen to illustrate spatial point patterns with a variety of features and methods available for analysing them.

We have mainly been involved in applications within ecology and, as Lawson remarks, many other scientific disciplines provide equally important areas of application for spatial point processes. Concerning spatial epidemiology, we are to some extent sceptical regarding the relevance of spatial point process modelling, as points often represent locations of residences which essentially fall on a discrete grid. We agree of course with Häggström and Grabarnik and Särkkä that existing knowledge concerning the underlying physics or biology should be used in the modelling. The ants nests and the Norwegian spruces examples do include some biological knowledge, which is used in the modelling of interactions. In the whales and the rain forest trees examples, our interest is focused on inference for the intensity or the intensity function, and we hence apply rather crude models to take into account the clustering – see also the more detailed comments below.

Rain forest trees: The clustering of the rain forest trees is treated as a nuisance phenomenon that we nevertheless need to take into account in the inference for the intensity function parameters. As Stoyan points out, quantifying the clustering and studying the biological reason for it is an equally important topic. We agree that the pair correlation function is more informative than the K-function, but for minimum contrast estimation, the use of the pair correlation function requires extra tuning parameters: a bandwidth for kernel estimation and a truncation near zero (for clustered point patterns, the kernel estimate of the pair correlation function is biased near the origin). The estimates of the K-function obtained assuming either stationarity or second-order reweighted stationarity do not deviate much for the rain forest trees, which explains the good accordance between Stoyans empirical pair correlation function and the pair correlation obtained for our fitted log Gaussian Cox process. Moreover, it seems that Stoyans geostatistical approach is to some extent similar to our log Gaussian Cox process modelling: Assume that conditional on a Gaussian field Ψ, $\log Z(u)$ (using

Stoyan's notation) is normal with conditional mean given by the right-hand side of (8). Then the difference to the log Gaussian Cox process approach is that 'Poisson sampling error' is replaced by normal measurement error (nugget effect), and the order of log and conditional expectation is interchanged. Furthermore, is the geostatistical approach, which involves specifying the disk radius R, really superior to the log Gaussian Cox process model?

Minke whales: In reply to Lawson and Schoenberg, notice that our goal is relatively modest, as we just wish to quantify clustering so as to assess the uncertainty of the whale intensity estimate. Hence, the cluster model is not intended to provide biological understanding of the whale population. The space–time reasoning behind our whale approach is the following: the space–time process of whale positions is assumed to possess an equilibrium distribution given by a spatial cluster model. Although it is not clear from the abridged presentation in the current paper, Waagepetersen & Schweder (2006) do not regard the whale sightings along different transect legs as observations of a single spatial realization of whale positions. Waagepetersen & Schweder (2006) indeed obtain the likelihood as a product of likelihoods evaluated for each transect leg separately assuming approximate independence between data observed along different transects at different times. Rather than using a parametric model for the clustering, one might consider estimating the inhomogeneous K-function, but such an estimate is very uncertain (Waagepetersen & Schweder, 2006) given the small number of observations and the very elongated shape of the components of the observation window. Bootstrapping the distribution of the intensity estimate seems difficult because of the inhomogeneity caused by varying observation conditions; perhaps it might be possible to use the thinned bootstrap procedure mentioned by Guan. Regarding Guttorps remark on survey design, note that evaluating the uncertainty of the simple moment based intensity estimate $\tilde{\lambda} = n / \int_W p(u)\,\mathrm{d}u$ (where n is the number of observed whales within the window W and p is the detection probability) only requires a specification of the second-order properties of the whale process. (A simulation study in Waagepetersen & Schweder (2006) did not demonstrate a notable loss of efficiency from using $\tilde{\lambda}$ instead of the maximum-likelihood estimate.)

Norwegian spruces: This example may seem a bit contrived, as the repulsion in the point pattern is to a large extent due to forest management, as pointed out by Penttinen. However, our Gibbs model still seems a rather nice marked point process approach to quantifying the spatial distribution of the trees. Penttinen remarks that the marks (stem diameters) appear to be independent when analysed by a more traditional approach based on summary statistics. On the other hand, our data analysis shows that the interaction parameter ψ in the conditional intensity (30) is significant. Hence, the conditional intensity of a mark does depend on the neighbouring marks.

Inhomogeneity: Baddeley, Hahn and Jensen *et al.* stress the importance of accounting for inhomogeneity. The notion of second-order reweighted stationarity seems both intuitively and mathematically appealing when dealing with inhomogeneous point processes. It is moreover general in the sense that it is implied by translation invariance of the pair correlation function, which is a common characteristic for spatial point processes. (Our more general definition of second-order reweighted stationarity in Baddeley *et al.* (2000) is given in terms of another common characteristic namely the intensity function.)

Penttinen and Stoyan remark that second-order stationarity may not be a valid assumption for the rain forest data, but it is nevertheless a much better approximation than stationarity (commonly assumed in the literature on spatial point processes). It may be even better with an improved model for the intensity function, where we may, for example, take into account the influence of trees of other species and consider additional covariates regarding soil properties.

Hahn and Jensen *et al.* mention alternative models of inhomogeneity (locally scaling, cluster process with inhomogeneous mother process) illustrated by instructive plots, and we encourage and pursue research into these interesting models and associated summary statistics. Cluster processes with an inhomogeneous mother process as illustrated in the left panel of Fig. B may be obtained as shot-noise Cox processes or generalized shot-noise Cox processes (Møller & Torrisi, 2005). The simple approach of modelling inhomogeneity by a Markov point process with an inhomogeneous first-order term (Ogata & Tanemura, 1986; Stoyan & Stoyan, 1998) may still be appealing to many statisticians, not least due to the SPATSTAT software. Berthelsen & Møller (2007) discuss non-parametric Bayesian inference for such inhomogenous Markov point processes.

Baddeley mentions some examples of 'singular' inhomogeneity, such as the concentration of points close to a line segment process. Further examples include Blackwell (2001), Blackwell & Møller (2003), and Skare *et al.* (2007), where the line segment process is given by the edges of a random (and possibly disturbed) Voronoi tessellation. These papers also demonstrate the usefulness of a Bayesian approach for a complex spatial model.

Cox and cluster processes: Guan discusses the very different interpretations of cluster models and log Gaussian Cox processes and the problem of disentangling various factors causing clustering. Assuming second order reweighted stationarity, he then describes a thinned block bootstrap procedure that avoids the specification of second-order properties of the point process when assessing parameter estimates obtained using the first order estimating function (47). The covariance matrix of the regression parameter estimates, on the other hand, only depends on the second-order properties of the point process. Hence, the inference concerning the regression parameter estimates is rather insensitive to whether a log Gaussian Cox process or a cluster model is specified, provided the fitted K-functions are similar. When likelihood-based methods are used, the specific choice of model is likely to be more critical, cf. the different inferences for the log Gaussian Cox process depending on whether a Bayesian or an estimating function approach is used (see also Guttorp's contribution).

Jensen *et al*. consider first a Levy driven Cox process which turns out to be a shot noise Cox process (with trend), and second a log Levy driven Cox process (LLCP) which in fact is a Cox process with the random intensity $\Lambda(u) = \Lambda_1(u)\Lambda_2(u)$, where Λ_1 is a log Gaussian Cox process and Λ_2 is an independent log shot noise process (so a log Gaussian Cox process is the special case where $\Lambda_2 \equiv 1$). We expect a more tractable class of Cox processes is obtained by replacing the log shot noise process for a LLCP by a shot noise process, since

$$\rho^{(n)}(u_1, \ldots, u_n) = \mathrm{E}[\Lambda_1(u_1) \cdots \Lambda_1(u_n)] \times \mathrm{E}[\Lambda_2(u_1) \cdots \Lambda_2(u_n)]$$

where the first expectation is well known for a log Gaussian process (Møller *et al.*, 1998), while the last expectation is much simpler to evaluate if Λ_2 is a shot noise process rather than a log shot noise process.

Jensen *et al.* conclude by suggesting a two-step estimation procedure, but we disagree that this is simulation free for the following reasons. For transformation inhomogeneous Markov point processes, maximizing $L_2(\hat{\alpha}, \omega)$ (using the notation in Jensen *et al.*) requires Markov chain Monte Carlo (MCMC) as discussed in Nielsen & Jensen (2004). For a Cox process, for example, a stationary log Gaussian Cox process with exponential covariance function $c(u, v) = \sigma^2 \exp(-||u - v||/\beta)$ and intensity $\alpha = \exp(\mu + \sigma^2/2)$, the first step of estimation gives $\hat{\alpha} = n(\mathbf{x})/|W|$. If we let $\omega = (\exp(-\sigma^2/2), \beta)$, we obtain

$$L_2(\hat{\alpha}, \omega) = \mathrm{E}_{\hat{\alpha}, \omega}\left[\exp\left(\int_W \left(\hat{\alpha} - \mathrm{e}^{\Psi(s)}\, \mathrm{d}s\right)\right) \prod_{i=1}^{n} \left(\mathrm{e}^{\Psi(x_i)}/\hat{\alpha}\right)\right]$$

where the expectation is with respect to a stationary Gaussian process $\mathbf{\Psi}$ with mean μ and covariance function c as specified by $(\hat{\alpha}, \omega)$. This expectation requires again MCMC, and apart from having one parameter less, maximization of $L_2(\hat{\alpha}, \omega)$ seems just as hard as finding the maximum-likelihood estimate (MLE) based on the original likelihood.

Gibbs and Markov models: The problem with fitting the interaction range parameter in a Markov model is, as Baddeley remarks, to a large extent due to non-differentiability. This excludes gradient based methods for optimizing the likelihood or pseudolikelihood function, and standard asymptotic arguments based on Taylor expansions are not valid.

We agree when Baddeley argues that Markov models may be needed to interpret interaction, and when Kendall notices that, if a Gibbs model is the equilibrium distribution of an (algorithmic) time-reversible process, it does not itself disqualify the model. As Kendall suggests it seems an interesting question to investigate whether spatial Gibbs model may arise as the marginal distribution of a simple non-homogeneous space–time process at a given time point.

We also agree with Billiot that determinantal and permanental point processes are promising models. These are defined by weighted determinants and permanents, and the special case of a usual determinant (the fermion point process) seems particularly tractable when studying Gibbsian properties (Georgii & Yoo, 2005).

Kendall notices that the attractive area-interaction point process defined on all of $\mathbb{R}^2$ exhibits phase transition; this issue is further discussed in Häggström *et al.* (1999). Kendall asks what diagnostics would indicate when phase transition phenomenon might be a potential problem for a general point process model. This is indeed a good question of relevance in physics, although for large point patterns in biology and many other areas, phase transition may not be an issue because of inhomogeneity.

Grabarnik and Särkkä discuss how hierarchical modelling can also be used to model asymmetric interactions between trees of different size classes. We agree that taking into account a hierarchical structure is important, but the hierarchical approach is more difficult when the classes of trees are not qualitatively but quantitatively defined. There is an issue of how to choose 'cut-points' and also the size of the trees may not reflect an ordering in time; for example, a small tree may have appeared earlier than larger trees, and from a time point of view one should then condition on the small tree.

Guttorp mentions that the intensity is often scale-dependent with, say, components of clustering, repulsion and non-stationarity; Guan adds a similar comment. Guttorp asks if we should overlay these different processes in the style of a generalized additive model (GAM). For example, consider the random intensity $\Lambda(u)$ in (12), which describes clustering around the points in the mother process $\mathbf{\Phi}$ and non-stationarity caused by the covariates. We may extend this to a random Papangelou conditional intensity with pairwise interaction

$$\lambda(u, \mathbf{x} \mid \mathbf{\Phi}) = \Lambda(u) \exp\left(\sum_{v \in \mathbf{x}} \theta \cdot t(\{u, v\})\right)$$

where θ is an interaction parameter (compare with (25)). Conditional on $\mathbf{\Phi}$, this defines an inhomogeneous Markov point process with density

$$f(\mathbf{x} \mid \mathbf{\Phi}) = \frac{1}{c_{\omega,\,\theta,\,\mathbf{\Phi}}} \left(\prod_{u \in \mathbf{x}} \Lambda(u)\right) \exp\left(\sum_{\{u,v\} \subseteq \mathbf{x}} \theta \cdot t(\{u, v\})\right)$$

where ω denotes the unknown parameters in (12). This complicated likelihood may either be treated by a MCMC MLE missing-data approach or probably more conveniently by a Bayesian MCMC approach (as the normalizing constant $c_{\omega,\theta,\mathbf{\Phi}}$ is intractable, the auxiliary variable method in Møller *et al.*, 2006, may be needed). For somewhat similar situations, see Berthelsen & Møller (2006, 2007).

Aggregation: Guttorp notices that data nowadays are collected on different scales, and Cressie asks how aggregated data modelled by a Markov random field (MRF) may be described by an underlying point process, for example, in epidemiological applications. This issue is also discussed in Møller (2003), but we are not aware of any satisfactory solution. Considering Poisson and shot-noise Gaussian Cox processes with a degenerate kernel (Brix, 1999) leads to a trivial MRF with no interaction. Considering any of the known Cox or Markov point process models would not lead to a common MRF but to a very complicated lattice process, which may be best analysed by a MCMC missing data approach.

Residuals: From both a theoretical and a practical point of view, the use of residuals for spatial point processes still needs development; cf. the comments by Penttinen and Schoenberg. We believe that residuals should play an important role in model assessment. For example, in Illian *et al.* (2007), L-functions indicate deviations from an inhomogeneous Poisson process, but they do not inform whether the deviations are due to 'random' clustering or a misspecified model for the intensity. A posterior predictive analysis of residuals, on the other hand, suggests that there are no systematic deviations from the specified intensity model. Regarding the rain forest data, a residual analysis in Guan (2007) discloses an extraordinarily dense cluster of trees not accounted for by the inhomogeneous cluster model depending on topographic covariates.

Computational methods: We agree with Baddeley that both simulation-demanding likelihood-based methodology and quick simulation-free approaches will continue to be of importance. As Guttorp remarks, estimating functions based on first or second-order properties are less efficient than likelihood-based estimation. However, the theoretical advantage of likelihood-based methods may be partly lost when in practice we approximate the likelihood using Monte Carlo methods and hence introduce a Monte Carlo error in the estimation procedure. In a simulation study, the estimating function (47) performed well, but it was not possible to quantify the loss of efficiency due to intractability of MLE within the settings of the simulation study.

Lawson points out that the availability of user-friendly software is pertinent for more widespread application of spatial point process methodology. In reply to this and a question by Schoenberg, for spatial Poisson and Markov point processes, SPATSTAT provides a very flexible software package, which, for example, allows for routine fitting of interactions between covariates and model selection in a manner similar to ordinary linear and generalized linear models. For Cox processes, the simple estimating function approach available for second-order reweighted stationary Cox processes can easily be carried out using SPATSTAT. Likelihood-based inference for Cox processes (Møller & Waagepetersen, 2003) is a much less developed area. We, for example, tried to apply the cluster process MCMC methodology in Waagepetersen & Schweder (2006) to the rain forest data but faced problems with extremely slow mixing of the MCMC chain. Hence, more research into universally applicable MCMC methods for cluster processes seems needed. For log Gaussian Cox processes, MCMC methods based on fast Fourier transforms seem to work reasonably well but user-friendly software (e.g. an R package) still needs to be written. The computational developments for log Gaussian Cox processes by Rue *et al.* (2007) are very welcome and we hope that their methodology will soon find its way into publicly available software. Rue *et al.* mention the danger of separately updating the Gaussian field and the covariance parameters, but while they consider intrinsic Gaussian random fields, we have not experienced similar problems in our applications involving second order stationary Gaussian fields.

In the context of maximum pseudo-likelihood estimation, Lawson remarks that numerical integration based on Delaunay triangles performs better than Dirichlet weights, which is one of the options available in SPATSTAT for evaluating the pseudolikelihood. In the context of inhomogeneous Poisson and cluster processes, Waagepetersen (2007) suggests to replace deterministic numerical integration schemes in SPATSTAT with Monte Carlo approximations. This allows evaluation of the parameter estimation error resulting from the approximation of the integral in the Poisson likelihood, and theoretical results indicate that Dirichlet weights may not be optimal in the case of smooth covariates.

Häggström adds a cautionary note regarding geometric ergodicity. Geometric ergodicity is a qualitative property, and even in the case of uniform ergodicity, the mixing time may be extremely large when simulating a spatial point process; cf. Møller (1999). Indeed, as Häggström stresses, quantitative results are needed, however, yet no useful quantitative results have been provided for spatial point processes; cf. appendix B in Møller (1999). This is one reason why perfect simulation is appealing, although ordinary MCMC methods including a detailed output analysis based on time series plots, estimated autocorrelations, etc. (e.g. Dellaportas & Roberts, 2003; Møller & Waagepetersen, 2003), play the prominent role in practice.

Miscellaneous: Our paper has essentially restricted attention to statistical theory and practice for analysing a single spatial point pattern. We completely agree with Baddeley, Häggström, Jensen *et al.*, Kendall, Mateu, Penttinen, and Schoenberg when they stress the importance of space–time and marked point processes; cf. sections 10.1 and 10.5.

Statistical theory for many spatial point patterns, spatial point patterns observed with noise, point patterns with multiple points, interpretation of summary statistics are still under developments, as remarked in section 10.1 and by Baddeley and Schoenberg. Concerning spatial point patterns observed with noise, it may be worth noticing that a Poisson (or Cox) process with i.i.d. disturbances of the points results in another Poisson (or Cox) process, while a Markov point process with i.i.d. disturbances of the points does not result in another Markov process.

Penttinen states that we ignore the window problem. Sections 2, 6.1, 6.2, 7.2 and 8.1 discuss this important issue, but perhaps we should had added more; see, for example, Møller and Waagepetersen (2003).

We agree with Heinrich who stresses the importance of research in large domain statistics e.g. for goodness-of-fit tests based on non-parametric summary statistics like the K-function. Large domain statistics also becomes important in connection with parametric inference for inhomogeneous point processes as in Waagepetersen & Guan (2007) who establish the joint asymptotic normality of parameter estimates obtained using respectively (47) and minimum contrast estimation as in example 13.

Mateu discusses multi-generation point processes evolving through discrete time. It is rather straightforward to construct such models and study them by simulation, but the models may be hard to analyse otherwise. One exception is spatial Hawkes processes, where some moment results exist (Brémaud *et al.*, 2005; Møller & Torrisi, 2007).

References

Baddeley, A. (2006). Validation of statistical models for spatial point patterns. In *SCMA IV: Statistical challenges in modern astronomy, 2006, proceedings* (eds J. Babu & E. Feigelson). To appear.

Baddeley, A., Bischof, L., Sintorn, I. & Turner, R. (2007). Analysing designed experiments where the response is a spatial point pattern. In preparation.

Baddeley, A., Møller, J. & Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statist. Neerlandica* **54**, 329–350.

Banerjee, S., Carlin, B. P. & Gelfand, A. E. (2004). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, Boca Raton, FL.

Barndorff-Nielsen, O. E. & Schmiegel, J. (2004). Levy based spatio-temporal modelling, with applications to turbulence. *Russian Math. Surveys* **59**, 65–90.

Bell, M. & Grunwald, G. (2004). Mixed models for the analysis of replicated spatial point patterns. *Biostatistics* **5**, 633–648.

Berthelsen, K. K. & Møller, J. (2006). Bayesian analysis of Markov point processes. In *Case studies in spatial point process modeling* (eds A. Baddeley, P. Gregori, J. Mateu, R. Stoica & D. Stoyan), 85–97. Springer Lecture Notes in Statistics **185**, Springer-Verlag, New York.

Berthelsen, K. K. & Møller, J. (2007). Non-parametric Bayesian inference for inhomogeneous Markov point processes. Technical Report R-2007-09, Department of Mathematical Sciences, Aalborg University. Available at http://www.math.aau.dk/~jm.

Bertin, E., Billiot, J. M. & Drouilhet, R. (1999). Existence of nearest neighbour spatial Gibbs models. *Adv. in Appl. Prob.* **31**, 895–909.

Besag, J. E. (1974). Spatial interactions and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36**, 192–225.

Besag, J. E., Milne, R. K. & Zachary, S. (1982). Point process limits of lattice processes. *J. Appl. Probab.* **19**, 210–216.

Blackwell, P. G. (2001). Bayesian inference for a random tessellation process. *Biometrics* **57**, 502–507.

Blackwell, P. G. & Møller, J. (2003). Bayesian analysis of deformed tessellation models. *Adv. in Appl. Prob.* **35**, 4–26.

Brémaud, P., Massoulié, L. & Ridolfi, A. (2005). Power spectra of random spike fields and related processes. *Adv. in Appl. Prob.* **37**, 1116–1146.

Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Adv. in Appl. Prob.* **31**, 929–953.

Chayes, J. T., Chayes, L. & Kotecký, R. (1995). The analysis of the Widom-Rowlinson model by stochastic geometric methods. *Comm. Math. Phys.* **172**, 551–569.

Clausius, R. (1858). Über die mittlere Länge der Wege, welche bei Molecularbewegung gasförmiger Körper von den einzelnen Molecülen zurück gelegt werden, nebst einigen anderen Bemerkungen über die mechanische Wärmetheorie. *Ann. Phys.* **105**, 239–258.

Comas, C. & Mateu, J. (2007a). Multi-generation space-time point processes. *Statistics*, to appear.

Comas, C. & Mateu, J. (2007b). Growing and reproducing particles evolving through space and time. *Metrika*, DOI: 10.2007/s 00184-007-0127-6, in press.

Cowling, A. (1998). Spatial methods for line transect surveys. *Biometrics* **54**, 828–839.

Cressie, N., Zhu, J., Baddeley, A. J. & Nair, N. G. (2000). Directed Markov point processes as limits of partially ordered Markov models. *Methodol. Comput. Appl. Probab.* **2**, 5–21.

Dellaportas, P. & Roberts, G. O. (2003). An introduction to MCMC. In *Spatial statistics and computational methods* (ed. J. Møller), 1–41. Lecture Notes in Statistics **173**, Springer-Verlag, New York.

Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*. Hodder Arnold, London.

Diggle, P. J., Mateu, J. & Clough, H. E. (2000). A comparison between parametric and non-parametric approaches to the analysis of replicated spatial point patterns. *Adv. in Appl. Prob.* **32**, 331–343.

Fleischer, F., Beil, M., Kazda, M. & Schmidt, V. (2006). Analysis of spatial point patterns in microscopic and macroscopic biological image data. In *Case studies in spatial point process modeling* (eds A. Baddeley, P. Gregori, J. Mateu, R. Stoica & D. Stoyan), 235–260. Springer Lecture Notes in Statistics **185**, Springer-Verlag, New York.

Gangnon, R. E. (2006). Impact of prior choice on local Bayes factors for cluster detection. *Stat. Med.* **25**, 883–895.

Georgii, H.-O. (1994). Large deviations and the equivalence of ensembles for Gibbsian particle system with superstable interaction. *Probab. Theory Related Fields* **99**, 171–195.

Georgii, H.-O. & Yoo, H. (2005). Conditional intensity and Gibbsianness of determinantal point processes. *J. Stat. Phys.* **118**, 55–84.

Grabarnik, P. & Särkkä, A. (2007). Modelling the spatial structure of forest stands by Gibbs point processes with hierarchical interactions. In preparation.

Guan, Y. (2007). Variance estimation for statistics computed from inhomogeneous spatial point processes. *J. Roy. Statist. Soc. Ser. B*, to appear.

Guan, Y. & Loh, J. M. (2007). A thinned block bootstrap procedure for modeling inhomogeneous spatial point patterns. *J. Amer. Statist. Assoc.*, to appear.

Häggström, O., van Lieshout, M. C. N. M. & Møller, J. (1999). Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. *Bernoulli* **5**, 641–658.

Hartvig, N. V. (2002). A stochastic geometry model for functional magnetic resonance images. *Scand. J. Statist.* **29**, 333–353.

Hedley, S. L. & Buckland, S. (2004). Spatial Models for Line Transect Sampling. *J. Agric. Biol. Environ. Stat.* **9**, 181–191.

Heinrich, L. (1991). Goodness-of-fit tests for the second moment function of a stationary multidimensional Poisson process. *Statistics* **22**, 245–268.

Heinrich, L. (2007). Asymptotic goodness-of-fit tests for point processes based on scaled empirical K-functions. In preparation.

Hellmund, G. (2005). Levy driven Cox processes – with a view to modelling tropical rain forest. Master thesis, Department of Mathematical Sciences, University of Aarhus.

Hellmund, G., Prokešová, M. & Jensen, E. B. V. (2007). Lévy driven Cox point processes. In preparation.

Illian, J. B., Møller, J. & Waagepetersen, R. (2007). Spatial point process analysis for a plant community with high biodiversity. *Environ. Ecol. Stat*. Conditionally accepted.

Jensen, E. B. V., Jonsdottir, K. Y. & Schmiegel, J. (2007). Spatio-temporal modelling with a view to biological growth. In *Statistics in spatio-temporal systems* (eds B. Finkenstadt, L. Held & V. Isham), 47–75. Monographs on Statistics and Applied Probability, Chapman & Hall/CRC, Boca Raton.

Jensen, E. B. V. & Nielsen, L. S. (2001). A review on inhomogeneous spatial point processes. In *Selected proceedings of the symposium on inference for stochastic processes* (eds I. V. Basawa, C. C. Heyde & R. L. Taylor), 297–318. IMS Lecture Notes **37**, Institute of Mathematical Statistics.

Jensen, E. B. V. & Thorarinsdottir, T. (2007). A spatio-temporal model for fMRI data – with a view to resting state networks. *Scand. J. Statist*. **34**, 587–614.

Jonsdottir, K. Y., Schmiegel, J. & Jensen, E. B. V. (2007). Levy based growth modelling. *Bernoulli*, to appear.

Kelly, F. P. & Ripley, B. D. (1976). A note on Strauss's model for clustering. *Biometrika* **6**, 357–360.

Kendall, W. S. & Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Adv. in Appl. Prob.* **32**, 844–865.

Kendall, W. S., van Lieshout, M. N. M. & Baddeley, A. J. (1999). Quermass interaction processes: conditions for stability. *Adv. in Appl. Prob.* **31**, 315–342.

Kingman, J. F. C. (1977). Remarks on the spatial distribution of a reproducing population. *J. Appl. Probab.* **14**, 577–583.

Lang, S. & Brezger, A. (2004). Bayesian P-splines. *J. Comput. Graph. Statist.* **13**, 183–212.

Lawson, A. B. (1992). GLIM and normalising constant models in spatial and directional data analysis. *Comput. Statist. Data Anal.* **13**, 331–348.

Lawson, A. B. (1993). A deviance residual for heterogeneous spatial Poisson processes. *Biometrics* **49**, 889–897.

Lawson, A. B. (2000). Cluster modelling of disease incidence via RJMCMC methods: a comparative evaluation. *Stat. Med.* **19**, 2361–2375.

Lawson, A. B. & Denison, D. (2002). *Spatial Cluster Modelling*. Chapman & Hall, London.

Lawson, A. B., Kulldorff, M., Simeon, S., Biggeri, A. & Magnani, C. (2007). Line and point cluster models for spatial health data. *Comput. Statist. Data Anal.*, to appear.

Marcelli, E. & Martinelli, F. (1996). Some new results on the kinetic Ising model in the phase coexistence region. *J. Stat. Phys.* **84**, 655–696.

Mase, S. (1992). Uniform LAN condition of planar Gibbsian point processes and optimality of maximum likelihood of soft-core potential functions. *Probab. Theory Related Fields* **92**, 51–67.

Mase, S. (1999). Marked Gibbs processes and asymptotic normality of maximum pseudo-likelihood estimators. *Math. Nachr.* **209**, 151–169.

Matérn, B. (1960). *Spatial variation*. Meddelanden från Statens Skogsforskningsinstitut **49**, 1–144. Statens Skogsforskningsinstitut, Stockholm.

Matérn, B. (1971). Doubly stochastic Poisson processes in the plane. In *Statistical ecology*, Vol. I (eds G. P. Patil, E. C. Pielou & W. E. Waters), 195–213. Pennsylvania State University Press, University Park.

McKeague, I. & Loiseaux, M. (2002). Perfect sampling for point process cluster modelling. In *Spatial cluster modelling* (eds A. B. Lawson & D. Denison), 87–108. CRC press, Boca Raton.

Møller, J. (1999). Markov chain Monte Carlo and spatial point processes. In *Stochastic geometry: likelihood and computation* (eds O. E. Barndorff-Nielsen, W. S. Kendall & M. N. M. van Lieshout), 141–172, Monographs on Statistics and Applied Probability **80**, Chapman and Hall/CRC, Boca Raton, FL.
Møller, J. (2003). A comparison of spatial point process models in epidemiological applications. In *Highly structured stochastic systems* (eds P. J. Green, N. L. Hjort & S. Richardson), 264–268. Oxford University Press, Oxford.
Møller, J., Pettitt, A. N., Berthelsen, K. K. & Reeves, R. W. (2006). An efficient MCMC method for distributions with intractable normalising constants. *Biometrika* **93**, 451–458.
Møller, J., Syversveen, A. R. & Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scand. J. Statist*. **25**, 451–482.
Møller, J. & Torrisi, G. L. (2005). Generalised shot noise Cox processes. *Adv. Appl. Probab.* **37**, 48–74.
Møller, J. & Torrisi, G. L. (2007). The pair correlation function of spatial Hawkes processes. *Statist. Probab. Lett*. **77**, 995–1003.
Møller, J. & Waagepetersen, R. (2003). *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC, Boca Raton, FL.
Monestiez, P., Sampson, P. D. & Guttorp, P. (1993). Modelling of heterogeneous spatial correlation structure by spatial deformation. *Cahiers de Geostatistique* **3**, 1–12.
Neyman, J. (1939). On a new class of 'contagious' distributions, applicable in entomology and bacteriology. *Ann. Math. Statist.* **10**, 35–57.
Nguyen, X. X. & Zessin, H. (1979). Ergodic theorems for spatial processes. *Z. Wahrsch. verw. Geb.* **48**, 133–158.
Nielsen, L. (2000). Modelling the position of cell profiles allowing for both inhomogeneity and interaction. *Image Anal. Stereol.* **19**, 183–187.
Nielsen, L. S. & Jensen, E. B. V. (2004). Statistical inference for transformation inhomogeneous Markov point processes. *Scand. J. Statist.* **31**, 131–142.
Ogata, Y. & Katsura, K. (1988). Likelihood analysis of spatial inhomogeneity for marked point patterns. *Ann. Inst. Statist. Math.* **40**, 29–39.
Ogata, Y. & Tanemura, M. (1986). Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns. In *Proceedings of the Pacific Statistical Congress* (eds I. S. Francis, B. F. J. Manly & F. C. Lam), 150–154. Elsevier, Amsterdam.
Prokešová, M., Hahn, U. & Jensen, E. B. V. (2006a). Statistics for locally scaled point processes. In *Case studies in spatial point process modelling* (eds A. Baddeley, P. Gregori, J. Mateu, R. Stoica & D. Stoyan), 99–123. Springer Lecture Notes in Statistics **185**, Springer, New York.
Prokešová, M., Hellmund, G. & Jensen, E. B. V. (2006b). On spatio-temporal Levy based Cox processes. In *Proceedings S4G, International Conference on Stereology, Spatial Statistics and Stochastic Geometry, Prague, June 26 to 29 2006* (eds R. Lechnerova, I. Saxl & V. Beneš), 111–116. Union Czech Mathematicians and Physicists, Prague.
Renshaw, E., Mateu, J. & Saura, F. (2007). Disentangling mark/point interaction in marked point processes. *Comput. Statist. Data Anal*. **51**, 3123–3144.
Rue, H. & Held, L. (2005). *Gaussian Markov random fields: theory and applications*. Monographs on Statistics and Applied Probability **104**, Chapman & Hall, London.
Rue, H., Martino, S. & Chopin, N. (2007). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. Technical report no. 1/2007, Department of Mathematical Sciences, Norwegian University of Science and Technology.
Ruelle, D. (1970). Superstable interactions in classical statistical mechanics. *Commun. Math. Phys.* **18**, 127–159.
Särkkä, A. & Renshaw, E. (2006). The analysis of marked point patterns evolving through space and time. *Comput. Statist. Data Anal.* **51**, 1698–1718.
Schoenberg, F. (2006). On non-simple marked point processes. *Ann. Inst. Statist. Math.* **58**, 223–233.
Shimatani, K. (2002). Point processes for fine-scale spatial genetics and molecular ecology. *Biom. J.* **44**, 325–352.
Skare, Ø., Møller, J. & Jensen, E. B. V. (2007). Bayesian analysis of spatial point processes in the neighbourhood of Voronoi networks. *Statistics and Computing*, in press.
Stein, M. L. (1999). *Interpolation of spatial data: some theory for Kriging*. Springer-Verlag, New York.
Stoyan, D. & Stoyan, H. (1995). *Fractals, random shapes and point fields*. Wiley, Chichester.
Stoyan, D. & Stoyan, H. (1998). Non-homogeneous Gibbs process models for forestry – a case study. *Biom. J.* **40**, 521–531.

Svedberg, T. (1922). Ett bidrag till de statistiska metodernas använding inom växtbiologi. *Svensk Botanisk Tidskrift* **16**, 1–8.
Taskinen, I. (2001). Cluster priors in the Bayesian modelling of fMRI data. PhD thesis, Department of Statistics, University of Jyväskylä, Jyväskylä.
Tscheschel, A. & Stoyan, D. (2006). Statistical reconstruction of random point patterns. *Comput. Statist. Data Anal.* **51**, 859–871.
Waagepetersen, R. (2005). Discussion of the paper by Baddeley, Turner, Møller & Hazelton (2005). *J. Roy. Stat. Soc. Ser. B* **67**, 662.
Waagepetersen, R. (2007). Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. Technical Report R-2007-04, Department of Mathematical Sciences, Alborg University.
Waagepetersen, R. & Guan, Y. (2007). Two-step estimation for inhomogeneous spatial point processes. In preparation.
Waagepetersen, R. & Schweder, T. (2006). Likelihood-based inference for clustered line transect data. *J. Agric. Biol. Environ. Stat.* **11**, 264–279.
Wang, Y. & Lawson, A. B. (2006). An evalaution of trianglation weights in fitting spatial Poisson process models. Submitted.
Widom, B. & Rowlinson, J. S. (1970). A new model for the study of liquid-vapor phase transitions. *Journal of Chemical Physics* **52**, 1670–1684.
Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99**, 250–261.