![Norsk Regnesentral — Norwegian Computing Center logo]
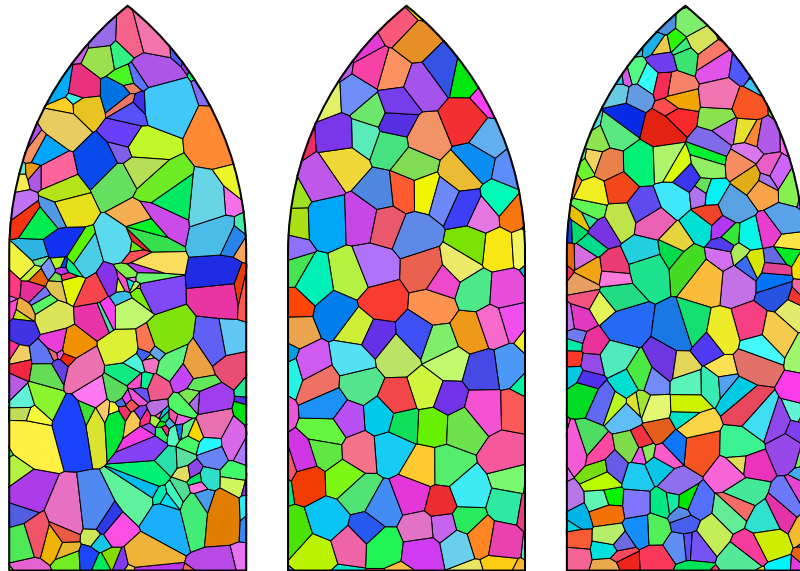
Bayesian Inference for Non-Markovian Point Processes

Note

Note no SAMBA/54/10

Authors Peter Guttorp
Thordis L. Thorarinsdottir

Date 23rd November 2010

The authors

Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modelling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

Title	**Bayesian Inference for Non-Markovian Point Processes**
Authors	**Peter Guttorp , Thordis L. Thorarinsdottir**
Date	23rd November 2010
Publication number	SAMBA/54/10

Abstract

The Bayesian approach to statistical inference has in recent years become very popular, especially in the analysis of complex data sets. This is largely due to the development of Markov chain Monte Carlo methods, which expand the scope of application of Bayesian methods considerably. In this paper, we review the Bayesian contributions to inference for point processes. We focus on non-Markovian processes, specifically Poisson and related models, doubly stochastic models, and cluster models. We also discuss Bayesian model selection for these models and give examples in which Bayes factors are applied both directly and indirectly through a reversible jump algorithm.

1 Introduction

Statistical inference for point processes originates, as pointed out by Daley and Vere-Jones (2005), in two sources: life tables, and counting phenomena. Among early sources of inferential work are Graunt, Halley and Newton in the 18th century on the life table side, and Newcomb, Abbé and Seidel in the second half of the 19th century on the counting side (for Newcomb's contributions, see Guttorp (2001); the others are all described by Daley and Vere-Jones). The modern approach originated mainly in England in the 1950s and 60s, with Bartlett and Cox as the main names. A few examples of point process patterns are shown in Figure 1.

This paper will review the Bayesian contributions to inference for point processes. We will only discuss non-Markovian processes, as lately much of the emphasis has been on Markovian models, and we consider it important not to lose sight of the non-Markovian ones. We make no pretense of a complete literature review; rather, we have chosen papers that we think are interesting or important or that we can use to make a point. A more comprehensive review paper is Møller and Waagepetersen (2007). Chapter 4 of the recent Handbook of Spatial Statistics (Gelfand et al., 2010) is devoted to spatial point processes.

Figure 1. Examples of point process patterns. Left: a process which is both clustered and regular (a Matérn type I process, section 2.4, yields the cluster centers and a Neyman-Scott process, section 4, the points). Middle: a regular process (a Matérn type I process). Right: completely random process (Poisson process, section 2.1). Shown is the Voronoi tesselation of the points. Generated by Ute Hahn and Dietrich Stoyan.

We start in section 2 by reviewing work on nonparametric estimation (Bayesian is always assumed unless otherwise specified) of the rate of a nonhomogeneous Poisson process. Immediately we will see that many processes, and many inference problems, can be viewed from more than one point of view. We then proceed to models derived from a Poisson process using thinning, and show how one can use Bayes factors to distinguish between models of late fall precipitation in upstate New York, USA.

The next section 3 deals with doubly stochastic models, and again we encounter the prob-

lem of how one views the inference. Section 4 deals with cluster processes, where we show an application to brain imaging, and section 5 is about model selection. Here we compare the Akaike criterion and the Bayes factor for selecting between types of cluster models. Finally, a short summary is given in section 6.

We are grateful to the organizers of the Toledo conference for the opportunity to participate and to write this paper. We also need to thank the Norwegian Computing Center in Oslo, Norway, and the University of Heidelberg, Germany, for accommodating visits by one or the other of us. Alex Lenkoski provided many helpful comments, and we thank Ute Hahn and Dietrich Stoyan for generating the patterns in Figure 1.

2 Poisson and related processes

2.1 Nonparametric estimation for nonstationary Poisson rate functions

In 1978 Aalen (1978) revolutionised point process analysis by introducing a general nonparametric statistical theory for the class of multiplicative counting processes. It was a frequentist theory, but received a Bayesian adaptation in the work of Lo (1982) for Poisson processes, and Lo and Weng (1989) for the general multiplicative processes. Kim (1999) also dealt with the general multiplicative process, but used a Lévy process prior. Here we will focus on Lo's treatment of the Poisson process case. Consider a Poisson process with intensity measure ν. Lo showed that a gamma process prior is conjugate. To define the gamma process prior, consider a σ-finite measure α, and say the measure μ is selected by a gamma process prior if for disjoint sets $A_1, ..., A_k$ we have that the collection of random variables $\{\mu(A_1), ..., \mu(A_k)\}$ are independent gamma random variables of scale 1 and means $\alpha(A_i)$. The measure μ is then said to have shape measure α and scale parameter 1. We denote the corresponding probability measure having these finite-dimensional distributions by $P_{\alpha,1}$. We can rescale the measure by an α-integrable positive random function β by defining $\beta\mu(A) = \int_A \beta(x)\mu(dx)$ and the corresponding probability measure is denoted $P_{\alpha,\beta}$. Lo showed that if we observe independent realisations $N_1, ..., N_n$ of N, and assign a prior measure $P_{\alpha,\beta}$ to the intensity measure ν, then the posterior measure is $P_{\alpha+\sum_1^n N_i, \beta/(1+n\beta)}$.

Consider now the special case where $\beta(x) = 1/\theta$, and suppose we are interested in estimating the intensity $\nu_t = \nu(0,t]$ under integrated squared error loss. It is not difficult to verify that the Bayes estimator is

$$\frac{\theta}{\theta+n}\frac{\alpha(0,t]}{\theta} + \frac{n}{n+\theta}\frac{1}{n}\sum_{i=1}^n N_i(0,t],$$

i.e., a weighted average of the prior guess and the sample empirical estimate.

Generally the tools needed to estimate nonparametrically a nonhomogeneous Poisson process with time dependent rate $\lambda(t)$, assumed integrable over the period of observation A, are the same as those for density estimation. Conditionally upon the total number of points $N = N(A)$ the points are distributed as the order statistics from a distribution with density (Cox and Lewis, 1966)

$$f(s) = \lambda(s)/\int_A \lambda(u)du. \tag{1}$$

Diggle (1985) used this fact to develop a kernel estimator for the intensity, and Peeling et al. (2007) used a histogram type estimator in setting up a Bayesian analysis of an interesting problem in musicology. In order to create a Bayesian structure, it has been popular to assign a prior related to a Gaussian process, typically of the form $\exp(X(t))$ where $X(t)$ is a Gaussian process. By the same misnomer as for the lognormal distribution, this tends to be called a log Gaussian Cox process, although it is the log intensity which is Gaussian, and in our context serves as a prior for a nonhomogeneous Poisson process intensity, while the setup mathematically (albeit not conceptually) corresponds to a doubly stochastic Poisson process (Cox, 1955). The doubly stochastic Poisson process if of course of interest in its own right (see section 3). The conditional likelihood for this model, given the realisation of $\lambda(s), s \in A$, is simply the usual Poisson likelihood

$$L(\lambda(s)) = \exp\Big(\int_A (\log \lambda(s)dN_s - \lambda(s)ds)\Big). \tag{2}$$

For random infinite dimensional $\lambda(s)$ the integral in the exponential of (2) cannot be evaluated explicitly, which makes Bayesian inference with a prior $Y(t)$ based on a Gaussian process intractable. Cressie and Rathbun (1994) and Møller et al. (1998) used a discretisation approach to obtain a tractable expression for the likelihood and Beneš et al. (2005) applied this to the Bayesian problem we are considering in this section. The idea is to approximate the continuous process $Y(t)$ by a sequence of step functions in the linear case, and values on a grid in the spatial case. Waagepetersen (2004) showed that the resulting posterior density converges to the true posterior as the discretisation interval shrinks to zero. Both he and Beneš et al. (2005) pointed out the sensitivity of the resulting inference to the discretisation scheme.

Heikkinen and Arjas (1998) took a similar route, using piecewise constant functions with random number of jumps of random size as prior on the intensity function, but not thinking of this as an approximation to a smooth prior process. It does not follow, for example, that the posterior mean is piecewise constant. In fact, it typically comes out smooth.

Kottas (2006) used the representation in (1) to develop a different estimation method. Treating $\gamma = \int_A \lambda(u)du$ as a separate parameter, he used explicit density estimation tools to estimate $f(s)$. We let $A = (0, T]$. Then f is estimated as a Dirichlet mixture of scaled beta densities (supported on $(0, T]$). The Dirichlet process is determined by a precision parameter α, which is given a gamma prior, and a base distribution, which is a function of the location and dispersion of the beta distributions. These are taken to be independent

uniform and inverse gamma, respectively. Finally, γ is given a Jeffreys prior of the form $1/x$.

2.2 The thinning approach to simulation

Lewis and Shedler (1979) introduced the standard approach to generating nonhomogeneous Poisson processes on a set A. If the rate is $\lambda(s)$ and we write $\lambda^* = \sup_{t \in A} \lambda(t)$, their *thinning* approach is to generate a homogeneous Poisson process of rate λ^*, and then keep a point at location τ with probability $\lambda(\tau)/\lambda^*$. This is, of course, a form of rejection sampling.

Adams et al. (2009) extended the Lewis-Shedler method to enable exact computation of the posterior distribution of a nonhomogeneous Poisson process with a Gaussian process prior of the form $\lambda^* \sigma(X(s))$, where $\sigma(x) = (1 + \exp(-x))^{-1}$, by keeping track of the deleted locations as well as the values of the Gaussian process at both the deleted and the kept locations, which they think of as a latent variable. Their approach is to use a Markov chain Monte Carlo approach containing three types of steps: changing the number of deleted points, the locations of the deleted points, and the values of the Gaussian process. The likelihood of this finite system can then be written down explicitly, without the need to evaluate integrals of Gaussian processes. Their approach appears to outperform the discretisation approach of the previous subsection on smooth intensity functions.

2.3 Extensions

Many of the methods for point processes on the line generalise to spatial processes. In some cases these extensions are non-straightforward, mainly concerning the lack of well-ordering of $\mathbb{R}^2$. A fairly recent review is Kottas and Sanso (2007), section 2.4. Interesting applications include Skare et al. (2007) who modelled a spatial pattern of badger territories and the distribution of pores in 3D translucent alumina using an inhomogeneous Poisson process with high intensity near the edges of an unobserved Voroni tessellation. We have chosen not to focus on parametric rate models (which abound e.g. in the software reliability literature, Huang and Bier (1999); Kuo and Yang (1996)), since most of these are very similar to Bayesian models for iid data.

2.4 Matérn thinning

Matérn (1960) introduced three different thinnings of Poisson processes in order to produce point processes that were more regular than the Poisson. Type I simply deletes all pairs of points that are within a radius R of another point. This is perhaps the simplest hard core rejection model in the literature. Type II introduces independent uniform marks t_i, called times, to each of the original points. The point with the smallest mark among all neighbours within distance R is retained. Clearly this model would have a higher rate of points than type I. Matérn also considered a third, dynamic variant, which Huber and

Wolpert (2009) called Type III, and which Matérn thought intractable. The retained points are called "seen", while the removed points are "hidden". In the type III process the seen points are those for which no seen point with lower time mark lies within distance R. So, for example, if we have three points with increasing times, such that the first is within R of the second, and the second is within R of the third, we have no points left in a type I process, only the first point left in the type II process, but potentially two points left in the type III process (see Huber and Wolpert (2009), Figure 1, for a graphical illustration).

In order to calculate the likelihood for a type III process, Huber and Wolpert used a technique akin to that used by Adams et al. in the previous subsection. Specifically, they suggested starting with n seen points and parameters λ and R, and then draw hidden points from a Poisson process of rate λ, and draw time marks uniformly for both seen and hidden points, until for all hidden points there is a seen point within distance R and with smaller time mark. This has the drawback that it can take quite a long time if there is a large number of seen points. Define the *shadow* of a seen point configuration as the union of balls of radius R centred at each seen point cross the interval $(t_i, 1]$ containing the possible hidden points. Let $d\Lambda(x, t)$ be the joint intensity of a Poisson point at x with mark t. Then the density (with respect to a Poisson process with uniform marks) of a seen point pattern $\mathbf{x}$ with marks $\mathbf{t}$ becomes

$$\mathbb{1}\big(\rho(\mathbf{x}) > R\big)\lambda^n \exp(|S|(1 - \lambda)) \exp(\Lambda(D(\mathbf{x}, \mathbf{t}))) \tag{3}$$

where $\rho(\mathbf{x}) = \min_{i \neq j}(x_i, x_j)$ is the smallest interpoint distance and $D(\mathbf{x}, \mathbf{t})$ is the shadow of $(\mathbf{x}, \mathbf{t})$. It is straightforward to verify that the acceptance-rejection approach outlined above samples directly from the likelihood. A faster perfect simulation approach was also outlined, and has been expanded upon in Møller et al. (2010).

The likelihood calculation can now form the basis for a Bayesian approach to estimating parameter of a Matérn type III process. To our knowledge this has not yet been implemented elsewhere in the literature.

Example 2.1. (Comparison of cluster processes for precipitation models)

Hobbs and Locatelli (1978) described mesoscale rainfall activity in cyclonic storms roughly as follows. Synoptic scale weather fronts contain large mesoscale regions, rainfall bands, where precipitation activity is possible. In turn, these bands contain moving rain cells, which are the points of higher rainfall rates. Observing this from a fixed point in space (e.g., a rain gauge), we see varying amounts of rainfall over time, with precipitation tending to come in clusters. Mathematically, Le Cam (1960) was first to suggest modeling rainfall at a location by a cluster point process, while Kavvas and Delleur (1981) proposed a Neyman-Scott Poisson cluster process, in which the primary process is a non-homogeneous Poisson process, and were the first to fit it to observed data. In a sequence of papers in the 1980's, a variety of cluster process approaches were developed (a review is
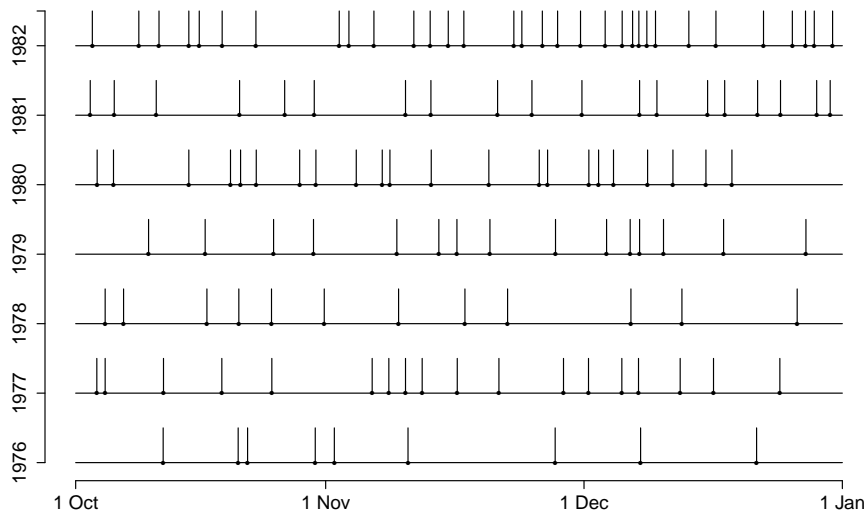
Figure 2. Fall (October through December) precipitation events observed at Whiteface Mountain, New York, 1976-1982.

provided in Guttorp, 1996; Salim and Pawitan, 2003, discusses more recent work), usually made stationary by considering only a short time period each year, such as a month.

In most versions of cluster point process analysis of precipitation, the primary process is assumed unobserved. This may be reasonable if only rain gauge data is used. However, one would often be able to assess the arrival of weather fronts using different types of data. Guttorp (1988) used so-called event-based data from the MAP3S acid rain monitoring network to assess features of the secondary process. This is the same data set that we will be using for our analysis, see Figure 2. The MAP3S/PCN (Multistate Atmospheric Power Product Pollution Study / Precipitation Chemistry Network) network of nine monitoring stations in the northeastern United States was initiated in 1976. We will focus on station 1, located on Whiteface Mountain in New York, at an altitude of 610 meters. The data were obtained from the Battelle-Pacific Northwest Laboratories ADS (Acid Deposition System) data base. They are described in Gentleman et al. (1985), and in MAP3S/RAINE Research Community (1982). The data were collected on an event basis, using samplers that open during precipitation, and close during dry periods. The definition of an event in the MAP3S network was left to the operator of the station; the Whiteface operator made a meteorologically based decision on what constitutes a new event.

For either station, each event may contain several precipitation incidents, indicated by separate lid openings. Since storm fronts do not arrive according to a Poisson process (since the fronts are physically separate), we do not expect a Poisson cluster process to be an adequate description of precipitation. We thus perform a comparison of a homogeneous Poisson cluster model and the type III Matérn model described in the previous subsection. Here, we view the data as seven independent realisations of fall precipitation events at

Whiteface Mountain.

In a Bayesian framework, Bayes factors (Jeffreys, 1935) offer a natural way of scoring models based on the evidence provided by the data. Specifically, suppose that $p(\mathbf{x}|\theta, M)$ is the density function of the observed point pattern $\mathbf{x}$ under model M given the model-specific parameter vector θ. Let the prior density of θ (assumed to be proper) be given by $\pi(\theta)$. The marginal likelihood of $\mathbf{x}$ under model M is given by

$$m(\mathbf{x}|M) = \int p(\mathbf{x}|\theta, M)\pi(\theta|M)d\theta. \tag{4}$$

Two models, M_1 and M_2 may then be compared by calculating the Bayes factor

$$\mathrm{B}_{12}(\mathbf{x}) = \frac{m(\mathbf{x}|M_1)}{m(\mathbf{x}|M_2)}. \tag{5}$$

For our data set, the Matérn Type III density in (3) becomes

$$p(\mathbf{x}|\lambda, R, M_{\mathrm{Ma}}) = \mathbb{1}\big(\rho(\mathbf{x}) > R\big)\lambda^n \exp(7T + \lambda(nR - 7T)),$$

where $n = 127$ is the total number of observed points, $T = 92$ is the number of days in the observation period, and $\rho(\mathbf{x}) = 0.75$ is the minimum interpoint distance over all the seven realisations. Similarly, the density for the homogeneous Poisson process is given by,

$$p(\mathbf{x}|\lambda, M_{\mathrm{Po}}) = \lambda^n \exp(7T(1 - \lambda)).$$

Here, we assume that $\lambda, R > 0$. We assign the parameter λ a conjugate prior density and set it to be exponential with rate parameter $\nu = 2$ in both models, while the parameter R in the Matérn Type III density is assigned a uniform prior on $(0, T)$, see Figure 4. The Bayes factor for equiprobable models then becomes

$$\mathrm{B}_{\mathrm{Ma,Po}}(\mathbf{x}) = \frac{(7T + \nu)}{Tn^2}\left(\left(\frac{7T + \nu}{7T + \nu - n\rho(\mathbf{x})}\right)^n - 1\right) = 273618, \tag{6}$$

which strongly favours the Matérn Type III model which is consistent with our hypothesis. As shown in Figure 3, the value of the Bayes factor is highly dependent on the value of the minimum interpoint distance $\rho(\mathbf{x})$.

Based on the results above, we continue with the analysis of the Matérn Type III model only. The full conditional posterior distribution for λ is given by a $\Gamma(n + 1, 7T + \nu - nR)$ distribution and

$$p(R|\mathbf{x}, \lambda, M_{\mathrm{Ma}}) = \mathbb{1}\big(0 < R < \rho(\mathbf{x})\big)\frac{\lambda n}{\exp(\lambda n\rho(\mathbf{x})) - 1}\exp(\lambda nR).$$

Figure 4 shows the posterior distributions for R and λ which are obtained from 50000 simulations using a Gibbs sampler and the inverse transform. The posterior distributions are much sharper than the prior distributions and the posterior means are very close to the maximum likelihood estimates. The maximum likelihood estimates are given by $\hat{R} = 0.75$ and $\hat{\lambda} = 0.23$, while we obtain the posterior means $\tilde{R} = 0.74$ and $\tilde{\lambda} = 0.23$.
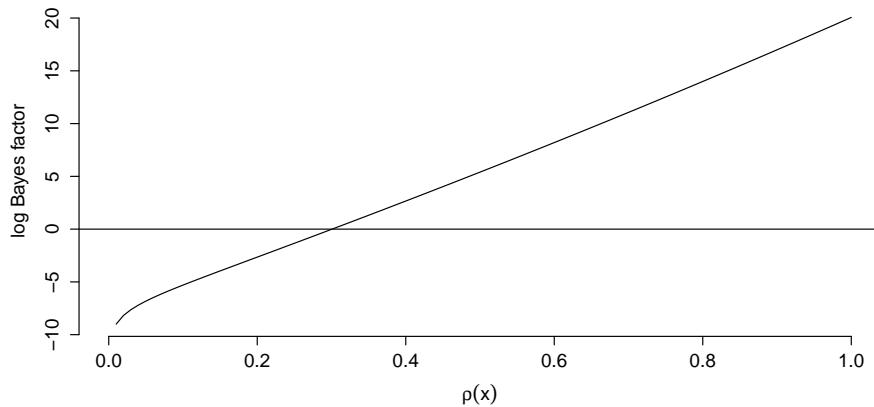
Figure 3. The Bayes factor for comparing the Matérn Type III model and the homogeneous Poisson model for the Whiteface Mountain precipitation data, as a function of the minimum interpoint distance, $\rho(\mathbf{x})$. The Bayes factors are plotted on a log-scale; values greater than zero favour the Matérn Type III model.
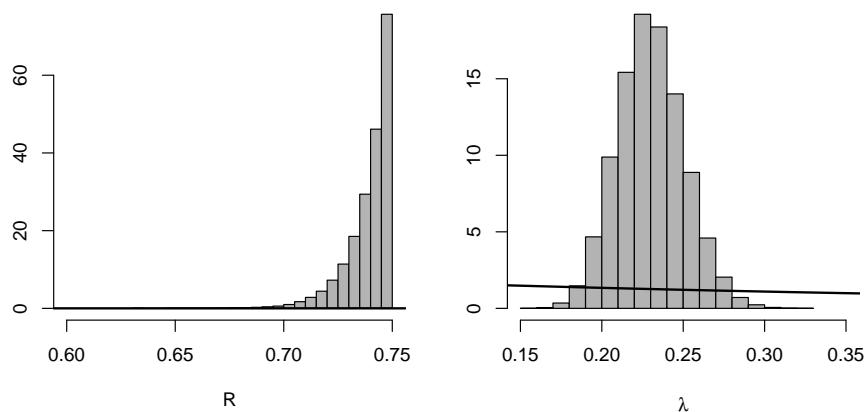

Figure 4. Posterior distributions for the parameters R (left) and λ (right) in the Matérn Type III model for precipitation events at Whiteface Mountain. The respective prior distributions are denoted by solid black lines.

3 Doubly stochastic processes

The doubly stochastic Poisson process, introduced by Cox (1955) and so named by Bartlett (1963) is obtained by letting the rate $\lambda(t)$ of the Poisson process vary according to a positive stochastic process, say $\Lambda(t)$. There are instances of doubly stochastic Poisson processes that are identical to cluster processes (for example, the shot noise process is a Neyman-Scott Poisson cluster process, see Daley and Vere-Jones (2005, p. 171-172). It is worth noting here that, except when the rate process is determined by the scientific situation, it is difficult to analyse a doubly stochastic process without having repeated observations, since the model is indistinguishable from a nonhomogeneous Poisson process based on a single path (see Møller and Waagepetersen, 2004). Thus, how you view your analysis can be seen as a matter of preference or convenience. We have not been able to find any Bayesian analyses of data where repeated observations are available, so that one can tell apart the doubly stochastic mechanism from the nonhomogenous Poisson process model.

Wolpert and Ickstadt (1998) modeled a spatial Poisson process with random intensity, where the intensity measure is a kernel mixture with a gamma measure. An an example, they analyse the density and spatial correlation of hickory trees. The same data were also analysed in Chapter 10.4 of Møller and Waagepetersen (2004) in a Bayesian setting using a nonhomogeneous Poisson process with a log-Gaussian prior process, where the Gaussian process has constant mean β, variance σ^2, and an exponential correlation function with range parameter α. The hyperparameters $\beta, \sigma^2, \kappa = \log(\alpha)$ need prior distributions as well. They used Jeffreys priors for the mean and variance, and a uniform prior between -2 and 4 for κ. As discussed in section 2, a discrete approximation to the prior process was used. The analysis was very sensitive to the prior on κ, and compared to a frequentist method of moment analysis using the g-function, the Bayesian method indicates a substantially larger correlation range. As pointed out above, this can also be viewed as a parametric Bayesian analysis of the doubly stochastic Poisson process obtained using a log Gaussian rate function.

Gutiérrez-Peña and Nieto-Barajas (2003) considered a doubly stochastic Poisson process with a gamma process (as in Lo, 1982) being its rate function $\Lambda(t)$. This process has parameter functions α (the rate function measure) and β (the scale process). In the case of constant scale $\beta = b$, the resulting process is what they call a negative binomial process of type 2. To perform a Bayesian analysis, they assigned a gamma process prior to the rate function measure α, and computed a closed form expression for the posterior distribution of α given the data. The authors did not view the distribution of the rate function $\Lambda(t)$ as a prior distribution.

Rue et al. (2009), in their highly influential paper on integrated nested Laplace approximations (INLA), illustrated how their numerical alternative to Markov chain Monte Carlo

methods can be applied to a doubly stochastic Poisson process where the intensity process is log Gaussian, although the method would work for any positive function of a Gaussian process such that the resulting doubly stochastic Poisson process is valid. The calculation of over 20,000 marginal distributions, applied to the rainforest data set also analysed by Waagepetersen (2007), took four hours of computing time. Again, the Gaussian process was discretized to a fine grid. To get similar precision with MCMC methods would be prohibitive computationally. It is possible within INLA to calculate Bayes factors, as noted in section 6.2 of the paper. However, the prior distributions used for the underlying random field are usually improper which means that the Bayes factor is only determined up to an unknown ratio of constants.

4 Cluster processes

The general cluster process (Daley and Vere-Jones, 2005) consists of a parent point process X to each point τ_i of which is associated a secondary point process $Z_i + \tau_i$. The structure of the primary and secondary processes is a suitable source of classification. Thus we can for example separate Poisson cluster processes (in which the primary process is Poisson) from general cluster processes (with a general primary process). On the line the most common secondary processes are of the Bartlett-Lewis type in which a random number of secondary points are laid out according to a renewal process, and the Neyman-Scott type where the secondary points are iid around the parent point (or cluster center). The named processes that abound in the literature (Cox cluster process, Matérn cluster process, Thomas process etc.) are simply special cases, and it does not seem useful to us to have a nomenclature which separates the particular distributional assumption. For example, we would call the Thomas process a Poisson cluster process of Neyman-Scott type using a Poisson cluster size distribution and normally distributed dispersion. Most Poisson cluster processes are non-Markovian; the exception being those with uniformly bounded cluster diameters (Baddeley et al., 1996).

Lieshout and Baddeley (2002) developed likelihood expressions for cluster processes with Poisson distributed offspring sizes, and developed Bayesian inference for processes where the prior distribution of the parent process is a Markov inhibition process. Of course, one could assign a Poisson process or a Matérn type III prior, and the results would be very similar. The main tool is a Markov chain Monte Carlo algorithm that uses a birth and death process (or, in a special case, coupling from the past), and the techniques are applied to a classical data set.

McKeague and Loizeaux (2002) considered Neyman-Scott processes in the plane, and also used an inhibition process as prior on the parent process. They used perfect sampling, and applied their tools to an example involving leukemia cases, where unobserved cluster

centers are estimated to lie close to some hazardous waste sites.

The idea of self-exciting processes is to have the rate depend on the development of the model in the past. If this dependence can be written as a linear functional, there is an alternative representation of this process as a cluster process (see Daley and Vere-Jones (2005), pp. 183-185). Gamerman (1992) used a variant where the intensity is piecewise constant, but dependent on the events in the previous piece. One could of course also think of this as a doubly stochastic model. Gamerman writes down equations for filtering and prediction as well as for Bayesian estimation of the rates in each interval.

Waagepetersen and Schweder (2006) used a Neyman-Scott process with negative binomial cluster size distribution and truncated bivariate normal dispersion to model minke whale populations. The data are obtained from line transect samples, and are modelled as a random thinning of the cluster process. The parameter of interest is the product of the rate of the cluster centers and the mean cluster size, called the whale intensity. It is estimated using Markov chain Monte Carlo, even though the exact likelihoods are computable.

Example 4.1. (Modelling activation in the human brain)

Functional magnetic resonance imaging (fMRI) is a technique for non-invasive *in vivo* recording of brain activation. It is based on the different magnetic properties of oxygenated and deoxygenated haemoglobin; images obtained with the method show changing blood flow in the brain associated with neural activation. Figure 5 shows such data set, where the subject was not exposed to stimuli during the recording of the data. Despite the lack of specific stimuli, changes in the signal appear over time, some of which show covariation in different regions of the brain.
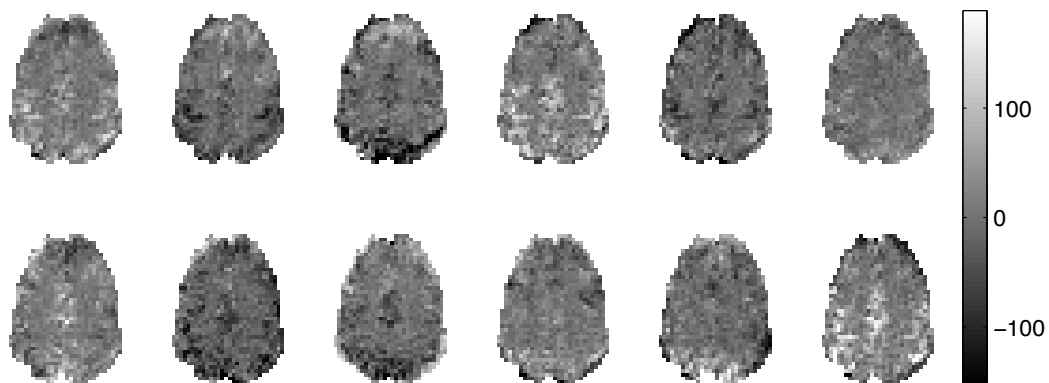

Figure 5. Development of the magnetic resonance (MR) signal activity over time in a single slice through the human brain. From left to right and top to bottom: the activity at time $t = 12, 30, 48, \ldots, 210$ seconds. Note that the images shown here have been preprocessed to correct for movement related artefacts and the signal changes have been enhanced so that they can be observed with the naked eye. From Thorarinsdottir and Jensen (2006).

In Thorarinsdottir and Jensen (2006) and Jensen and Thorarinsdottir (2007), a Bayesian spatio-temporal point process model for such data was proposed. Purely spatial processes for this type of data have also been proposed in Taskinen (2001) and Hartvig (2002). The activation is described by a marked point process Φ, where the point process is latent and corresponds to the unobserved neural activation while the marks are observed and describe the associated observed MR signal changes due to changes in the blood oxygenation level. It is thus the latent point process, Ψ, and the associated intensity function that are of main interest for the statistical analysis.

Assume that we have observed data $\{Z_{tx}\}$, where $t \in [0, T]$ denotes time and $x \in \mathcal{X}$ denotes a spatial location, or a voxel, in the brain region $\mathcal{X}$ which is a bounded subset of $\mathbb{R}^2$ or $\mathbb{R}^3$. Here, $\mathcal{X}$ is a single slice through the brain, $\mathcal{X} \subset \mathbb{R}^2$. To account for edge effects in the time domain, we assume that Ψ is a process on $[T_-, T] \times \mathcal{X}$, where $T_- < 0$ is chosen such that it is very unlikely that a neural activation starting before time T_- will affect an observed MR signal on $[0, T]$. The marked process is denoted by $\Phi = \{[t_i, x_i; m_i]\}$ with $(t_i, x_i) \in \Psi$ and marks $m_i \in \mathbb{R}^d$.

The resulting model for the observed MR signal intensity at time t and voxel x becomes

$$Z_{tx} = \mu_x + \sum_i f_{tx}(t_i, x_i; m_i) + \sigma \varepsilon_{tx}, \tag{7}$$

where μ_x is the baseline signal at voxel x and ε_{tx} is an error term with mean 0 and variance 1. The function f_{tx} describes the contribution to the observed signal intensity at voxel x and time t caused by a neuronal activation at $(t_i, x_i) \in \Psi$. This function is assumed to be separable in space and time with $f_{tx}(t_i, x_i; m_i) = g(t - t_i; m_i)h(x - x_i; m_i)$ and $m_i = (\theta_{1i}, \theta_{2i}, \theta_{3i}) \in \mathbb{R}^3_+$, where

$$h(y; m) = \theta_1 \exp\left(-\frac{\|y\|^2}{2\theta_2} \right)$$

and

$$g(u; m) = \int_0^{\theta_3} \frac{1}{\sqrt{2\pi}3} \exp\left(-\frac{(u - v - 6)^2}{18} \right) dv.$$

Here, $\| \cdot \|$ donotes the Euklidian norm. The mark parameters thus have the following interpretation: θ_{1i} describes the magnitude of the signal change due to neural activation i, θ_{2i} describes the spatial extend of this change, and θ_{3i} its temporal duration.

For simplicity, assume that the marks m_i and the variance σ^2 are fixed. The aim of the statistical analysis is then to recover the latent point process Ψ and its intensity function based on the observations $\{Z_{tx}\}$ under the model described by (7). Further, we may replace Z_{tx} in (7) by $Z_{tx} - \bar{Z}_{\cdot x}$ and f_{tx} by $f_{tx} - \bar{f}_{\cdot x}$. The new data have $\mu_x = 0$ and the same correlation structure as the original data if the number of observed images is sufficiently large.

The prior distribution of Ψ is chosen as Poisson with intensity λ. There is thus no interaction between points in the prior distribution and any interactions found in the posterior

distribution derive from interactions observed in the likelihood. The intensity function λ is assumed to be of the following form

$$\lambda(t, x) = \sum_{k=1}^{K} \lambda_k \mathbb{1}\big(x \in \mathcal{X}_k\big), \tag{8}$$

where the sets $\mathcal{X}_k \subseteq \mathcal{X}$ are disjoint. Their union may be the whole observed brain region $\mathcal{X}$ but need not be. The sets $\mathcal{X}_k$ should be specified by the experimenter while the parameters λ_k are unknown. The intensity function can be written as $\lambda(t, x) = c\lambda_2(x)$, where $c > 0$ and $\int_{\mathcal{X}} \lambda_2(x)dx = 1$. If follows from (8), that λ_2 can be written as

$$\lambda_2(x) = \sum_{k=1}^{K} \pi_k \frac{\mathbb{1}\big(x \in \mathcal{X}_k\big)}{|\mathcal{X}_k|},$$

where $|\cdot|$ denotes area and $\pi_k > 0$ with $\sum_{k=1}^{K} \pi_k = 1$. The parameters $c, \pi = (\pi_1, \ldots, \pi_K)$ are assigned non-informative prior distributions.

The posterior density is of the form

$$p(c, \pi, \psi | z) \propto p(z|\psi)p(\psi|c, \pi)p(c)p(\pi),$$

where the likelihood is given by

$$p(z|\psi) = [2\pi\sigma^2]^{-n(z)/2} \exp\Big(-\frac{1}{2\sigma^2} \sum_{t,x} \Big[z_{tx} - \sum_{(t_i, x_i) \in \psi} f_{tx}(t_i, x_i; m) \Big]^2 \Big).$$

A fixed scan Metropolis within Gibbs algorithm is used to simulate from the posterior density where in each scan c, π, and ψ are updated in turn. The full conditional distributions for c and π are given by a Gamma and a Dirichlet distribution, respectively, while the full conditional distribution for ψ is

$$p(\psi|c, \pi, z) \propto c^{n(\psi)} \prod_{k=1}^{K} \pi_k^{n_k(\psi)} \exp\Big(-\frac{1}{2\sigma^2} \sum_{t,x} \Big[z_{tx} - \sum_{(t_i, x_i) \in \psi} f_{tx}(t_i, x_i; m) \Big]^2 \Big), \tag{9}$$

where $n_k(\psi)$ denotes the number of points in ψ that fall within $\mathcal{X}_k$. Note that the full conditional distribution for ψ is in fact a pairwise interaction density. The point process ψ is simulated from the density in (9) using a Birth-Death-Move algorithm as described in Møller and Waagepetersen (2004).

Based on an earlier analysis of the same data set by Beckmann et al. (2005), the prior intensity in (8) was set to be positive in three sub-regions of interest, the left and right motor cortex and a middle region. The resulting posterior spatial intensity pattern for ψ when cumulated over time is shown in Figure 6. The posterior spatial intensity is clearly inhomogeneous in contrast to the homogeneous prior intensity with strong indications for clustering in the spatial domain.
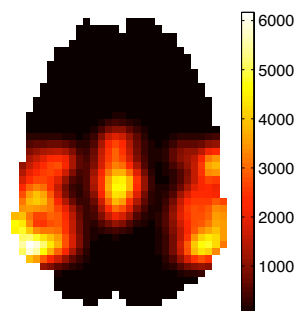
Figure 6. The posterior spatial activation pattern in the three regions of interest cumulated over time. The three regions are the left and right motor cortex and a middle region. From Jensen and Thorarinsdottir (2007).

5 Model selection

Model selection for point process models is commonly carried out by investigating the summary statistics of the point pattern prior to the model fitting. Formal Monte Carlo tests of goodness-of-fit to the homogeneous Poisson process or comparison of the nearest-neighbour distance distribution function and the spherical contact distribution function can provide the modeller with evidence for regularity or clustering in the point pattern as compared to complete randomness (Baddeley, 2010; Illian et al., 2008). Such comparisons can produce important guidance for choosing the correct class of models, yet these model classes are very broad, rendering the information less valuable.

Statistical inference for point process models is usually very computationally intensive, and it is often not feasible to perform inference for a single data set under many different models. For this reason, scientific understanding of the data, combined with expert knowledge of the model class, is often combined to *a priori* select a single model for a given data set, once the appropriate class of models has been established. However, if the scientific question of interest relates to specific details in the modelling, such as particularities in the clustering mechanism of the point pattern, a more formal procedure for model comparison is called for.

The Akaike information criterion (AIC), which is given by

$$\text{AIC} = -2 \log L + 2k, \tag{10}$$

where L is the maximum likelihood value and k is the number of parameters in the model, is by far the most popular model comparison criterion used in the point process literature. The AIC has the advantage that it can be applied to any likelihood based inference method. However, it has been noted that it tends to favour more complicated models for larger data sets (Ogata, 1999). This is a clear disadvantage in a setting where the modelling easily becomes computationally intractable. We discuss this issue further in Example 5.1.

Bayes factors (see Example 2.1) were first used in a point process context by Akman and Raftery (1986), who compared parametric intensity models for nonhomogenous Poisson processes on the line. The focus of this work was to develop conditions for which the Bayes factor could be determined under vague prior information. In this context, Akman and Raftery call the Bayes factor $B_{12}^{(n)}(\mathbf{x}, T)$ *operational* if for $U_n(T) = \{u = (u_1, \ldots, u_n) : 0 \le u_1 \le \ldots \le u_n \le T\}$, there exists a positive integer n such that

$$\sup_{T>0} \sup_{u \in U_n(T)} B_{12}^{(n)}(u, T) < \infty.$$

Then, m, the smallest such integer, is the smallest number of observed events needed for a comparison of M_1 and M_2. Furthermore, if $B_{12}^{(n)}(u, T)$ is a bounded function of u for each fixed n and T, and invariant to scale changes in the time variable,

$$B_{12}^{(n)}(u, T) = B_{12}^{(n)}(au, aT) \quad \forall a > 0,$$

for all n, u, T, then the Bayes factor is operational. It is thus, under fairly general conditions, sufficient to define the prior distributions such that the Bayes factor becomes time-invariant for it to be well defined. Akman and Raftery demonstrated this explicitly for log-polynomial intensity models.

Walsh and Raftery (2005) used partial Bayes factors for hypothesis testing to classify a point pattern as either a homogeneous Poisson pattern or a mixture of a homogeneous Poisson pattern and a hard-core Strauss process. Here, the term *partial Bayes factor* refers to calculating the Bayes factor in (5) using a summary statistic $\mathbf{y}$ rather than the full data $\mathbf{x}$, as the marginal likelihood is intractable for the mixture model considered in the study. The partial Bayes factor is equivalent to (5) if and only if $\mathbf{y}$ is a sufficient statistic for $\mathbf{x}$ under both M_1 and M_2.

To our knowledge, the work by Akman and Raftery (1986) and Walsh and Raftery (2005), are the only applications of Bayesian model selection criteria reported in the literature in the context of point process models as those, discussed in this paper. In Example 2.1, we showed how Bayes factors may be calculated directly for simple models. In the following example, we consider using a reversible jump algorithm for Bayesian model selection when direct calculation of the Bayes factor is not feasible.

Example 5.1. (Model selection for point processes of Neyman-Scott type)

Here, we compare using AIC and Bayes factors for model selection within the class of Neyman-Scott cluster processes. More precisely, we compare two different models of Neyman-Scott type which differ in the dispersion process for the secondary points. Model M_1 has a homogeneous Poisson cluster process, a Poisson cluster size distribution, and the dispersion distribution is given by a normal distribution. Model M_2, on the other hand, can be seen as a mixture of two such processes, where the dispersion variance differs for

the two components of the mixture. Palm likelihood inference for M_1 and M_2 was considered in Tanaka et al. (2007), where the two models are called the Thomas model and generalized Thomas model of type B, respectively. Prokešová and Jensen (2010) showed that the Palm likelihood estimator for these models is consistent and asymptotically normally distributed.

Model M_1 has a latent cluster centre process and three unknown parameters: the intensity of the cluster process, κ, the mean cluster size, α, and the dispersion variance, ω^2. We generate ten samples from this model on $B = [0,1] \times [0,1]$ for $(\kappa, \alpha, \omega) = (50, 30, 0.03)$ and perform Palm likelihood inference and Bayesian inference for each sample under both model M_1 and M_2. The procedure is then repeated with data samples generated from model M_2. Model M_2 has two latent cluster centre processes and five unknown parameters. We set the true parameters as $(\kappa_1, \kappa_2, \alpha, \omega_1, \omega_2) = (25, 25, 30, 0.02, 0.04)$, where κ_i is the intensity and ω_i^2 is the dispersion variance for cluster process $i = 1, 2$. Examples of such point patterns are shown in Figure 7.
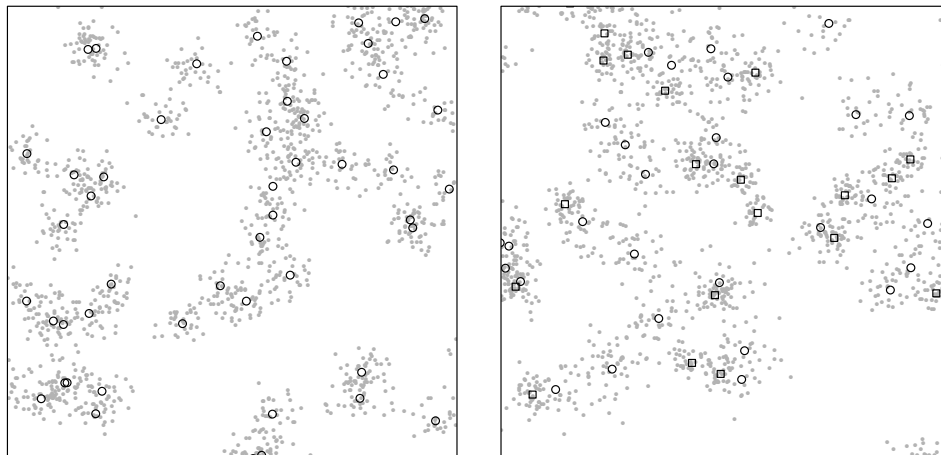

Figure 7. Examples of simulated point patterns of Neyman-Scott type in the plane. Left: Poisson cluster centre process with Poisson number of offsprings and normal dispersion process. Right: mixture of two such processes which differ in the dispersion variance. The observed point patterns are indicated with gray dots, while the black squares and circles indicate the latent cluster centre processes.

Bayesian inference for similar processes is discussed in e.g. Møller and Waagepetersen (2004), Møller and Waagepetersen (2007), and Waagepetersen and Schweder (2006). Contrary to the models considered in Example 2.1, we cannot calculate the marginal likelihood (4) of a dataset $\mathbf{x}$ under the models M_1 and M_2 directly. Instead, we define a reversible jump algorithm (Green, 1995) where we jump between the models M_1 and M_2. The Bayes factor can then be obtained directly from the MCMC sample by comparing the time spent in M_1 and the time spent M_2.

The random intensity function of M_2 is given by

$$\alpha Z(\xi|\Psi,\omega) = \alpha \left[\frac{1}{2\pi\omega_1^2} \sum_{c\in\Psi_1} \exp\left(-\frac{\|c-\xi\|^2}{2\omega_1^2}\right) + \frac{1}{2\pi\omega_2^2} \sum_{c\in\Psi_2} \exp\left(-\frac{\|c-\xi\|^2}{2\omega_2^2}\right) \right],$$

where $\omega = (\omega_1, \omega_2)$, $\Psi = (\Psi_1, \Psi_2)$ denotes the cluster centre processes, and $\|\cdot\|$ is the Euklidian norm. To account for edge effects, we define the centre processes on the extended window $B_{\text{ext}} = [-0.1, 1.1] \times [-0.1, 1.1]$. The density of a Poisson process on B with intensity function κ with respect to a homogeneous Poisson process X_1 with intensity λ is given by

$$p(\mathbf{x}|\kappa) = \exp\left(\lambda|B| - \int_B \kappa(\xi)d\xi\right) \prod_{\xi\in\mathbf{x}} \kappa(\xi),$$

where $|\cdot|$ denotes area. As noted by Møller and Waagepetersen (2004, p. 151), the choice of X_1 is not important for maximum likelihood inference and for MCMC simulations from a single model. However, when performing a reversible jump step between models with different number of latent processes, we need to choose λ with care in order to obtain balanced proposals, see below.

The joint posterior distribution of the latent processes and the parameters in M_2 is thus given by

$$p(\psi, \kappa, \alpha, \omega|\mathbf{x}) \propto p(\mathbf{x}|\alpha Z(\cdot|\psi,\omega))p(\psi_1|\kappa_1)p(\psi_2|\kappa_2)p(\kappa)p(\alpha)p(\omega),$$

and the joint posterior distribution under M_1 is an obvious simplification. Our MCMC simulation algorithm consists of the following steps:

(a) updating the latent process ψ;
(b) updating the parameter κ;
(c) updating the parameter α;
(d) updating the parameter ω;
(e) proposing to jump between M_1 and M_2.

Steps (a)-(d) are repeated 25 times under the same model between proposals to jump between models. For step (a), we use the Birth-Death-Move algorithm described in Møller and Waagepetersen (2004). If we are currently in model M_2, we propose one change for each of the latent processes ψ_1 and ψ_2. We assign conjugate priors to the parameters κ and α which result in closed form full conditional distributions for these parameters. More precisely, we set $\kappa \sim \Gamma(50,1)$, $\kappa_1, \kappa_2 \sim \Gamma(12.5, 0.5)$ and $\alpha \sim \Gamma(30,1)$, where the gamma distributions are parameterized in terms of shape and rate. The full conditional distributions are then

$$\kappa \mid \psi \sim \Gamma(50 + n(\psi), 1 + |B_{\text{ext}}|)$$

$$\kappa_i \mid \psi_i \sim \Gamma(12.5 + n(\psi_i), 0.5 + |B_{\text{ext}}|), \quad \text{for } i = 1, 2$$

$$\alpha \mid \mathbf{x}, Z(\cdot|\psi,\omega) \sim \Gamma\left(30 + n(\mathbf{x}), 1 + \int_B Z(\xi|\psi,\omega)d\xi\right).$$

A Metropolis-Hastings step is needed to update the dispersion parameter ω. We define the prior distribution for ω in terms of the precision and set $1/\omega^2 \sim \Gamma(1, 0.001)$. Under model M_2, we simulate initial values for ω_1 and ω_2 from the prior distribution until $\omega_1 < \omega_2$. This is needed for identifiability, as M_2 is otherwise invariant to permutations of the labels $i = 1, 2$. The joint prior distribution of (ω_1, ω_2) is thus 2 times the product of the individual prior components; this plays a role in the reversible jump step (e). To update the dispersion parameter ω under M_1, we generate a proposal $1/\omega^{2*} \sim \Gamma(1/\omega^2, 1)$ and accept it with probability

$$\min\left\{ \frac{p(\mathbf{x}|\alpha Z(\cdot|\psi, \omega^*)q(\omega|\omega^*)}{p(\mathbf{x}|\alpha Z(\cdot|\psi, \omega)q(\omega^*|\omega)}, 1 \right\},$$

where $q(\omega^*|\omega)$ is the proposal density for ω^* given the current state of the chain. Under M_2, the parameters ω_1 and ω_2 are updated in a similar way. However, a proposal is rejected immediately if the condition $\omega_1 < \omega_2$ is violated by the proposal.

The reversible jump step (e) is similar to the reversible jump step for normal mixtures described in Richardson and Green (1997). To move from M_2 to M_1 we need to merge the two cluster processes into one process. This is proposed by setting

$$\psi^* = \psi_1 \cup \psi_2$$
$$\kappa^* = \kappa_1 + \kappa_2$$
$$\omega^* = \sqrt{\frac{\kappa_1 \omega_1^2 + \kappa_2 \omega_2^2}{\kappa_1 + \kappa_2}}.$$

The reversible split move from M_1 to M_2 is now largely determined. There are two degrees of freedom involved in the split which we determine with a two-dimensional random vector u given by

$$u_1 \sim \text{Beta}(2, 2), \quad u_2 \sim \text{Beta}(2, 2).$$

Here, we set

$$\kappa^* = (\kappa_1^*, \kappa_2^*) = (u_1 \kappa, (1 - u_1)\kappa), \tag{11}$$
$$\omega^* = (\omega_1^*, \omega_2^*) = \left(\sqrt{\frac{u_2}{u_1}}\omega, \sqrt{\frac{1 - u_2}{1 - u_1}}\omega \right),$$

and reject the proposal immediately if $\omega_1^* < \omega_2^*$ does not hold. It still remains to allocate the points in ψ to either ψ_1^* or ψ_2^*. This is performed by allocating each point in ψ at random to either ψ_1^* with probability κ_1^*/κ or to ψ_2^* with probability κ_2^*/κ.

The acceptance probability for a split move is

$$\min\left\{ \frac{p(\psi^*, \kappa^*, \alpha, \omega^*|\mathbf{x})}{p(\psi, \kappa, \alpha, \omega|\mathbf{x})q(u)}|J|, 1 \right\},$$

where $q(u)$ is the density function of u and J is the Jacobian of the transformation described in (11),

$$|J| = \frac{\omega\kappa}{2\sqrt{u_1(1 - u_1)}\sqrt{u_2(1 - u_2)}}.$$

As mentioned above, we need to choose the densities of the latent cluster processes carefully in order to obtain balanced proposals. Let X, X_1, and X_2 be homogeneous Poisson processes on B_{ext} with intensities λ, λ_1, and λ_2, respectively, such that

$$\lambda = \frac{n(\psi)}{|B_{\text{ext}}|}, \quad \lambda_1 = \frac{n(\psi_1)}{|B_{\text{ext}}|}, \quad \lambda_2 = \frac{n(\psi_2)}{|B_{\text{ext}}|}.$$

The log-ratio of the density of (Ψ_1^*, Ψ_2^*) with respect to (X_1, X_2) and the density of Ψ with respect to X is then given by

$$\log\left(\frac{p(\psi_1^*|\kappa_1^*)p(\psi_2^*|\kappa_2^*)}{p(\psi|\kappa)}\right) = n(\psi_1^*)\left[\log\frac{\kappa_1^*}{\kappa} - \log\frac{n(\psi_1^*)}{n(\psi)}\right] + n(\psi_2^*)\left[\log\frac{\kappa_2^*}{\kappa} - \log\frac{n(\psi_2^*)}{n(\psi)}\right],$$

which penalizes for a lack of balance between the proposed intensites and the corresponding point patterns. The acceptance probability for a merge move is calculated in a similar fashion. The algorithm was implemented in R (R Development Core Team, 2009).

The Palm likelihood inference is performed as described in Tanaka et al. (2007), where the maximization is repeated five times for each sample using different starting values each time. We found that this was necessary, as different starting values would often give different results. The AIC in (10) is then calculated for each sample based on the optimal result obtained over the five runs. The MCMC chain is run for 300000 iterations over the steps (a)-(d). We assessed the convergence by running several such chains for each data set which give nearly identical results. The starting values for both inference methods are set as

$$\kappa \sim \text{Po}(50), \quad \alpha \sim \text{Po}(30), \quad \frac{1}{\omega^2} \sim \Gamma(1, 0.001),$$

under M_1 and similar under M_2. For the Bayesian inference, the initial latent centre processes are simulated from a Poisson model and the chain is started randomly in either M_1 or M_2. The Palm likelihood inference takes about 30-40 minutes on a standard desktop computer for a single data set. Running one MCMC chain takes about 1.5-2 hours on the same computer.

Table 1. Model selection results for comparing M_1 and M_2 based on Akaike information criterion (AIC) and Bayes factors (BF) for simulated data. The table reports the classification results for each of the model selection criteria based on ten simulated data sets from each model.

Correct model	AIC		BF	
	M_1	M_2	M_1	M_2
M_1	8	2	10	0
M_2	3	7	0	10

The results of the simulation study are reported in Table 1. In the Bayesian framework, all the MCMC chains would initially jump back and forth between the two models and

then settle in the correct model. Under M_1, this initial burn-in period was very short, or only about 5000 iterations. However, the mixing was slower under M_2, and about 100.000 iterations were needed before all the chains would settle in M_2. In the frequentist framework, a data set would be classified as belonging to either M_1 or M_2 based on the minimum AIC obtained for that data set. As Table 1 shows, 25% of the data sets were wrongly classified by this method. We did, however, not find any indications of the AIC preferring either the simpler or the more complicated model. Generally, though, we would obtain a much greater difference between the two AIC scores when M_2 was chosen as the correct model.

6 Summary

Statistical inference for point process models was initially performed in a frequentist manner, with the earliest work on Bayesian inference being published about three decades ago. In this paper, we have reviewed the Bayesian contributions for non-Markovian processes. Our aim was not to provide a complete literature review; rather, we have chosen to focus on those papers that we find especially important or interesting. In particular, we have tried to emphasize the variety of applications to which non-Markovian point process models have been applied to.

We have further emphasized the use of Bayesian methodology for model selection. We show how Bayes factors can be used to determine model probabilities for simple models without performing a full inference under each model. For more complicated models, this is usually no longer the case. In an example, we show how a reversible jump algorithm can be used to determine model probabilities when the marginal likelihoods for the competing models cannot be computed directly. Traditionally, model selection methods for point processes mainly aim at detecting repulsion or clustering in the point pattern and there seems to be a lack of methods that apply beyond this initial distinction. The results presented here suggest that Bayesian methodology might be applied to fill this gap, although further research is needed.

References

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics*, 6(4):701–726.

Adams, R. P., Murray, I., and MacKay, D. J. C. (2009). Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *Proceedings of the 26th International Conference on Machine Learning*, Montreal, Canada.

Akman, V. E. and Raftery, A. E. (1986). Bayes factor for non-homogeneous Poisson processes with vague prior information. *Journal of the Royal Statistical Society, Series B*, 48(3):322–329.

Baddeley, A. (2010). Modelling strategies. In *Handbook of Spatial Statistics*. Chapmann & Hall/CRC.

Baddeley, A., Van Lieshout, M., and Møller, J. (1996). Markov properties of cluster processes. *Advances in Applied Probability*, 28(2):346–355.

Bartlett, M. (1963). The spectral analysis of point processes. *Journal of the Royal Statistical Society, Series B*, 25(2):264–296.

Beckmann, C. F., Deluca, M., Devlin, J. T., and Smith, S. M. (2005). Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society B*, 360:1001–1013.

Beneš, V., Bodlák, K., Møller, J., and Waagepetersen, R. (2005). A case study on point process modelling in disease mapping. *Image Analysis and Sterology*, 24:159–168.

Cox, D. R. (1955). Some statistical methods connected with series of events (with discussion). *Journal of the Royal Statistical Society, Series B*, 17:129–164.

Cox, D. R. and Lewis, P. A. (1966). *Statistical Analysis of Series of Events*. Methuen, London.

Cressie, N. and Rathbun, S. L. (1994). Asymptotic properties of estimators for the parameters of spatial inhomogeneous poisson point processes. *Advances in Applied Probability*, 26:124–154.

Daley, D. and Vere-Jones, D. (2005). *An Introduction to the Theory of Point Processes*, volume Volume I: Elementary Theory and Methods. Springer, 2nd edition edition.

Diggle, P. (1985). A kernel method for smoothing point process data. *Applied Statistics*, 34:138–147.

Gamerman, D. (1992). A dynamic approach to the statistical analysis of point processes. *Biometrika*, 79(1):39–50.

Gelfand, A. E., DIggle, P., Fuentes, M., and Guttorp, P. (2010). *Handbook of spatial statistics*. Chapman & Hall/CRC.

Gentleman, R., Zidek, J. V., and Olsen, A. R. (1985). Map3s/pcn acid rain precipitation monitoring data base. Technical Report 19, Department of Statistics, University of British Columbia, Vancouver.

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732.

Gutiérrez-Peña, E. and Nieto-Barajas, L. E. (2003). Bayesian nonparametric inference for mixed Poisson processes. In Bernado, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., editors, *Bayesian Statistics 7*, pages 163–179. Oxford University Press.

Guttorp, P. (1988). Analysis of event-based precipitation data with a view toward modeling. *Water Resources Research*, 24(1):35–43.

Guttorp, P. (1996). Stochastic modeling of rainfall. In Wheeler, M. F., editor, *Environmental Studies: Mathematical, Computational, and Statistical Analysis*, pages 171–187. Springer, New York.

Guttorp, P. (2001). Simon Newcomb–astronomer and statistician. In Heyde, C. C. and Seneta, E., editors, *Statisticians of the Centuries*, pages 197–199. New York: Springer.

Hartvig, N. V. (2002). A stochastic geometry model for functional magnetic resonance images. *Scandinavian Journal of Statistics*, 29:333–353.

Heikkinen, J. and Arjas, E. (1998). Non-parametric Bayesian estimation of a spatial Poisson intensity. *Scandinavian Journal of Statistics*, 25(3):435–450.

Hobbs, P. V. and Locatelli, J. D. (1978). Rainbands, precipitation cores and generating cells in a cyclonic storm. *Journal of Atmospheric Science*, 35:230–241.

Huang, Y.-S. and Bier, V. M. (1999). A natural conjugate prior for the nonhomogeneous Poisson process with an exponential intensity function. *Communications in Statistics–Theory and Methods*, 28(6):1479–1509.

Huber, M. L. and Wolpert, R. L. (2009). Likelihood-based inference for Matérn Type-III repulsive point processes. *Advances in Applied Probability*, 41:958–977.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. Wiley.

Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. *Proceedings of the Cambridge Philisophical Society*, 31:203–222.

Jensen, E. B. V. and Thorarinsdottir, T. L. (2007). A spatio-temporal model for functional magnetic resonance imaging data - with a view to resting state networks. *Scandinavian Journal of Statistics*, 34:587–614.

Kavvas, M. L. and Delleur, J. W. (1981). A stochastic cluster model for daily rainfall sequences. *Water Resources Research*, 17:1151–1160.

Kim, Y. (1999). Nonparametric Bayesian estimators for counting processes. *Annals of Statistics*, 27(2):562–588.

Kottas, A. (2006). Dirichlet process mixtures of beta distributions, with applications to density and intensity estimation. In *Proceedings of the Workshop on Learning with Nonparametric Bayesian Methods*.

Kottas, A. and Sanso, B. (2007). Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *Journal of Statistical Planning and Inference*, 137(10, Sp. Iss. SI):3151–3163.

Kuo, L. and Yang, T. (1996). Bayesian computation for nonhomogeneous Poisson processes in software reliability. *Journal of the American Statistical*, 91(434):763–773.

Le Cam, L. M. (1960). A stochastic description of precipitation. In Neyman, J., editor, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 3*, pages 165–186, California, Berkeley.

Lewis, P. A. W. and Shedler, G. S. (1979). Simulation of a nonhomogeneous Poisson process by thinning. *Naval Logistics Quarterly*, 26:403–413.

Lieshout, M. N. M. and Baddeley, A. J. (2002). Extrapolating and interpolating spatial patterns. In Lawson, A. B. and Denison, D., editors, *Spatial Cluster Modelling*, pages 61–86. Chapman & Hall/CRC, Boca Raton, Florida.

Lo, A. Y. (1982). Bayesian nonparametric statistical inference for Poisson point process. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 59:55–66.

Lo, A. Y. and Weng, C.-S. (1989). On a class of Bayesian nonparametric estimates: Ii. hazard rate estimates. *Annals of the Institute of Statistical Mathematics*, 41(2):227–245.

MAP3S/RAINE Research Community (1982). The MAP3S/RAINE precipitation chemistry network: statistical overview for the period 1976-1979. *Atmospheric Environment*, 16:1603–1631.

Matérn, B. (1960). *Spatial variation*, volume 49 of *Meddelanden från Statens Skogsforskningsinstitut*. Statens Skogsforskningsinstitut, Stockholm.

McKeague, I. W. and Loizeaux, M. (2002). Perfect sampling for point process cluster modelling. In Lawson, A. B. and Denison, D., editors, *Spatial Cluster Modelling*, pages 87–107. Chapman & Hall/CRC, Boca Raton, Florida.

Møller, J., Huber, M. L., and Wolpert, R. L. (2010). Perfect simulation and moment properties for the Matérn type III process. *Stochastic Processes and their Applications*, 120:2142 – 2158.

Møller, J., Syversveen, A. R., and Waagepetersen, R. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25:451–482.

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

Møller, J. and Waagepetersen, R. P. (2007). Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, 34(4):643–684.

Ogata, Y. (1999). Seismicity analysis through point-process modeling: a review. *Pure and Applied Geophysics*, 155:471–507.

Peeling, P., Li, C.-f., and Godsill, S. (2007). Poisson point process modeling for polyphonic music transcription. *Journal of the Acoustical Society of America*, 121(4):EL168–EL175.

Prokešová, M. and Jensen, E. B. V. (2010). Asymptotic Palm likelihood theory for stationary point processes. Thiele Research Report nr. 2/2010, Department of Mathematical Sciences, University of Aarhus.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from: `http://www.R-project.org`.

Richardson, S. and Green, P. J. (1997). On Baeysian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B*, 59(4):731–792.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B*, 71:319–392.

Salim, A. and Pawitan, Y. (2003). Extensions of the bartlett-lewis model for rainfall processes. *Statistical Modelling*, 3:79–98.

Skare, Ø., Møller, J., and Jensen, E. B. V. (2007). Bayesian analysis of spatial point processes in the neighbourhood of Voronoi networks. *Statistical Computing*, 17:369–379.

Tanaka, U., Ogata, Y., and Stoyan, D. (2007). Parameter estimation and model selection for Neyman-Scott point processes. *Biometrical Journal*, 49:1–15.

Taskinen, I. (2001). *Cluster priors in the Bayesian modelling of fMRI data*. PhD thesis, Department of Statistics, University of Jyväskylä, Jyväskylä.

Thorarinsdottir, T. L. and Jensen, E. B. V. (2006). Modelling resting state networks in the human brain. In Lechnerová, R., Saxl, I., and Beneš, V., editors, *Proceedings S⁴G: International Conference on Stereology, Spatial Statistics and Stochastic Geometry*.

Waagepetersen, R. (2004). Convergence of posteriors for discretized log Gaussian Cox processes. *Statistics & Probability Letters*, 66(3):229–235.

Waagepetersen, R. and Schweder, T. (2006). Likelihood-based inference for clustered line transect data. *Journal of Agricultural, Biological, and Environmental Statistics*, 11(3):264–279.

Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous neyman-scott processes. *Biometrics*, 63:252Ð258.

Walsh, D. C. and Raftery, A. E. (2005). Classification of mixtures of spatial point processes via partial Bayes factors. *Journal of Compuational and Graphical Statistics*, 14:139–154.

Wolpert, R. L. and Ickstadt, K. (1998). Poisson/Gamma random field models for spatial statistics. *Biometrika*, 85(2):251–267.