

# Bayesian smoothing in the estimation of the pair potential function of Gibbs point processes

JUHA HEIKKINEN<sup>1</sup> and ANTTI PENTTINEN<sup>2</sup>

<sup>1</sup>*Finnish Forest Research Institute, Unioninkatu 40 A, FIN-00170 Helsinki, Finland.*

*E-mail: juha.heikkinen@metla.fi*

<sup>2</sup>*University of Jyväskylä, Department of Statistics, P.O. Box 35, FIN-40351 Jyväskylä, Finland.*

*E-mail: penttine@jyu.fi*

A flexible Bayesian method is suggested for pair potential estimation with a high-dimensional parameter space. The method is based on a Bayesian smoothing technique, commonly applied in statistical image analysis. For the calculation of the posterior mode estimator a new Monte Carlo algorithm is developed. The method is illustrated through examples with both real and simulated data, and its extension into truly nonparametric pair potential estimation is discussed.

*Keywords:* Bayesian smoothing; Gibbs processes; posterior mode estimator; Markov chain Monte Carlo methods; Marquardt algorithm; pair potential function

## 1. Introduction

Gibbs point processes are a natural class of models for point patterns exhibiting interactions between the points. By far the most widely applied form in practical analyses is that of pairwise interaction, where the scale and strength of interaction between two points are determined by a so-called pair potential function. For a stationary and isotropic process the pair potential is a function of the distance between the two points.

The estimation of Gibbs processes is notoriously difficult because of their analytically and numerically intractable normalizing functions. A summary of some of the methods developed for an approximate likelihood estimation can be found in Diggle *et al.* (1994). Geyer (1998) gives a comprehensive and highly recommendable exposition of the likelihood inference using Markov chain Monte Carlo (MCMC).

Nonparametric estimation has been largely ignored by researchers. One exception is the suggestion to use the nonparametric estimation of the pair correlation function and its approximate relation to the pair potential through the Percus–Yevic equation (Diggle *et al.* 1987; Fiksel 1988). The approximation is a result of a cluster expansion method, and it is

accurate only for sparse data. Van Lieshout and Baddeley (1996) introduced a nonparametric measure of the strength and range of interaction, but its statistical properties seem to be somewhat problematic.

This paper introduces a method which can be viewed as the first step towards a truly nonparametric Bayesian estimation of Gibbs processes with pairwise interactions. The pair potential is approximated by a step function having a large number of fixed jump points. The induced high dimension of the parameter space causes two kinds of problem. First, each component of the sufficient statistic is typically a function of a small number of point locations, which causes instability in the estimation. Second, the computational complexity increases rapidly with the dimension.

To combat the first problem we apply Bayesian smoothing by choosing a Markov chain prior which penalizes large differences between nearby values of the pair potential function. This idea originates in Bayesian image analysis; see Besag (1986).

As regards the computational complexity, we have found the full posterior analysis to be too demanding with the currently available machinery. Consequently, we have concentrated on the task of locating the posterior mode, which is computationally equivalent to that of finding the maximum likelihood estimate (MLE). Starting from the Monte Carlo Newton–Raphson algorithm of Penttinen (1984) and the Monte Carlo likelihood approach of Geyer and Thompson (1992), we arrived at an efficient algorithm by modifying the former into an MCMC approximation of the Marquardt algorithm (Marquardt 1963) and then combining the two. The first approximation to the posterior mode is obtained using the Monte Carlo Marquardt algorithm, where the first two differentials of the log-posterior are approximated by MCMC as in Penttinen (1984) and the final estimate is calculated using the Monte Carlo likelihood approximation. (The naming conventions applied here were introduced by Geyer 1998.)

Our approach of an approximating step function and a smoothing prior was motivated by the ideas in Arjas and Gasbarra (1994). Dealing with a tractable likelihood, they were able to perform a full posterior analysis over the space of all step functions, where also the number and locations of the jump points were variable. Further work along these lines can be found in Green (1995), Arjas and Heikkinen (1997), and Heikkinen and Arjas (1998; 1999). Denison *et al.* (1998) used sequences of piecewise polynomials instead of our step functions. Although we had to compromise by fixing the jump points and restricting to the posterior mode estimation, the rapid development of the MCMC methodology and the increasing computing power available may soon offer the tools for truly nonparametric pair potential estimation in the spirit of these authors.

The rest of this paper is organized as follows. In Section 2 we give a short introduction to Gibbs point processes and to some Monte Carlo methods for their estimation (Section 2.2). Section 2 also includes a more specific discussion of the problems associated with a high-dimensional parameter space (in Section 2.2.3). Section 3 describes our new approach, with the smoothing prior built in Section 3.1 and the hybrid algorithm for posterior mode estimation detailed in Section 3.2. The amacrine cells data, presented in Diggle and Gratton (1984), are analysed in Section 4 using the proposed technique. In addition, the result is compared with other solutions as summarized in Särkkä (1993), and a related simulation study is reported.

## 2. Gibbs point processes

Our method has been developed for the analysis of mapped point patterns  $\mathbf{x} = \{x_1, \dots, x_n\} \subset E \subset \mathbb{R}^d$  having a finite number  $n = n(\mathbf{x})$  of points in a bounded sampling window  $E$ . Let  $\Omega_n, n > 0$ , be the set of the configurations with  $n$  points, and let  $\Omega_0 = \{\emptyset\}$ . Then the space of finite point patterns on  $E$  is  $\Omega = \cup_{n=0}^{\infty} \Omega_n$ .

A point process on  $E$  can be defined as a random variable on the exponential space  $(\Omega, \mathcal{B}, \mu)$  over the measure space  $(E, \mathcal{A}, \nu)$ , where  $\mathcal{A}$  is the Borel  $\sigma$ -field of  $E$  and  $\nu$  is the Lebesgue measure on  $E$  (Carter and Prenter 1972). The  $\sigma$ -algebra  $\mathcal{B}$  is a family of sets  $B \subset \Omega$  such that  $B \cap \Omega_n$  is an element of  $\mathcal{A}^n$ . The probability measure  $\mu$  given by

$$\mu(B) = e^{-\nu(E)} \left[ \mathbf{1}(\emptyset \in B) + \sum_{n=1}^{\infty} \left\{ \frac{1}{n!} \int_{E^n} \mathbf{1}(\{x_1, \dots, x_n\} \in B) \prod_{i=1}^n \nu(dx_i) \right\} \right] \tag{2.1}$$

is the distribution of the homogeneous *Poisson process* with unit intensity, that is, the number  $n(\mathbf{x} \cap A)$  of points in  $A \in \mathcal{A}$  has the Poisson  $\{\nu(A)\}$  distribution and the counts in disjoint sets are independent. The Poisson process provides a model for point patterns with no interaction in the sense that locations of a subset of points do not contain any information on the locations of the rest.

In general, a *Gibbs point process* can be any random variable on  $(\Omega, \mathcal{B}, \mu)$  having a density  $f$  with respect to  $\mu$ . Normally this density is assumed to satisfy the hereditary condition:  $f(\mathbf{x}) > 0$  implies  $f(\mathbf{x}') > 0$  for all  $\mathbf{x}' \subset \mathbf{x}$ . The model is usually specified via an *energy function*  $U: \Omega \rightarrow \mathbb{R} \cup \{\infty\}$  so that

$$f(\mathbf{x}) = Z^{-1} \exp\{-U(\mathbf{x})\}, \tag{2.2}$$

where the *partition function*  $Z = \int_{\Omega} e^{-U(\mathbf{x})} \mu(d\mathbf{x})$  (a function of  $U$ ) normalizes  $f$  into a density.

### 2.1. Pairwise interaction processes

A stationary and isotropic process with *pairwise interaction* is defined by letting

$$U(\mathbf{x}) = \alpha n(\mathbf{x}) + \sum_{j < k} \phi(\|x_j - x_k\|), \tag{2.3}$$

where the parameter  $\alpha$  is known as the *chemical activity*, and  $\phi: (0, \infty) \rightarrow \mathbb{R} \cup \{\infty\}$  is the *pair potential function* which has the following interpretation. Values  $\phi(r) > 0$  indicate lower probability density for inter-point distances  $r$  than under the Poisson process; there is *repulsion* (or *inhibition*) at distance  $r$ . If  $\phi(r) < 0$ , we say that there is *attraction* at distance  $r$ . Usually a finite *range of interaction*,  $R$ , is assumed such that  $\phi(r) = 0$  for  $r > R$ . If  $\phi(r) > 0$  for  $r \leq R$ , then typical realizations will be more or less *regular* compared to a completely random arrangement.

For illustration and interpretation it is convenient to transform the pair potential function  $\phi$  into the *interaction function*  $h = \exp(-\phi)$ . This way the infinities transform to zeros and

$h(r)$  is roughly the likelihood of the occurrence of a pair of points at distance  $r$  from each other.

Some constraints are needed for  $\alpha$  and  $\phi$  in order for  $\exp(-U)$  to be integrable. In particular, negative (or partly negative) potentials may lead to improper or unstable distributions (Kelly and Ripley 1976; Gates and Westcott 1986), which means that the pairwise interaction processes are not very useful models for clustering. Things become much simpler if we condition on the event  $n(\mathbf{x}) = n$  for some fixed  $n$ . This eliminates the parameter  $\alpha$ , and we obtain a process with density

$$f_\phi(\mathbf{x}) = Z(\phi)^{-1} \exp \left\{ - \sum_{j < k} \phi(\|x_j - x_k\|) \right\} \quad (2.4)$$

with respect to the  $n$ -fold product  $\nu^n$  of  $\nu$ , a measure on  $\Omega_n$ . The normalizing function  $Z(\phi) = \int_{\Omega_n} \exp\{-\sum_{j < k} \phi(\|x_j - x_k\|)\} \nu^n(d\mathbf{x})$  is now finite for all pair potentials  $\phi : (0, \infty) \rightarrow \mathbb{R} \cup \{\infty\}$ . When modelling an observed pattern  $\mathbf{x}$ ,  $n$  is naturally chosen to be the observed number of points  $n(\mathbf{x})$ .

Even these conditional models seem to fail in exhibiting moderate clustering (Geyer and Thompson 1995; Møller 1998; Geyer 1998). Thus the models for clustering should be built on some other basis. One possibility is to use the nearest-neighbour Markov point processes (Baddeley and Møller (1989); see Baddeley and van Lieshout (1995), Baddeley *et al.* (1996) and Møller 1998) for some recent work along these lines.

In most applications only a part of the whole pattern is observed, and the unobserved points outside  $E$  interact with those of  $\mathbf{x}$ . Then it is more appropriate to consider conditional distributions of the form (2.4) conditioned on the points outside  $E$  (see Stoyan *et al.* 1995, Section 5.5.3), and account for the unobserved part by some kind of edge correction. A review of them is given in Ripley (1988, Chapter 2); for an alternative approach, see Geyer (1998).

## 2.2. Monte Carlo estimation

In general, it is impossible to evaluate the partition function  $Z$  by analytic or numerical integration, which makes the exact likelihood (or posterior) analysis infeasible. The idea of Monte Carlo estimation is to express the required quantities as expectations of functionals of  $\mathbf{x}$ , and approximate these by the corresponding averages in a sample simulated from  $f$ . While producing independent realizations of a Gibbs point process is usually either impossible or very slow, we can always apply MCMC to simulate dependent realizations of a Markov chain whose stationary distribution is  $f$ . Good reviews of MCMC in general are given in Tierney (1994), Besag *et al.* (1995), Green (1995), and Geyer (1998); related methods for simulating point processes can also be found in Preston (1977), Ripley (1979), Stoyan *et al.* (1995), Geyer and Møller (1994) and Møller (1998).

Processes conditioned to have a fixed number of points can be simulated by the original Metropolis method (Metropolis *et al.* 1953). In the version used in this study a sequence  $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(T)}$  of realizations is produced by starting from an arbitrary pattern  $\mathbf{x}^{(0)}$ , and in the  $t$ th step performing the transition from  $\mathbf{x}^{(t-1)}$  to  $\mathbf{x}^{(t)}$  as follows. Select an index

$i = i_t \in \{1, \dots, n\}$  randomly. Choose a uniform random point  $z_i$  from  $E$ , and set  $z_j = x_j^{(t-1)}$  for  $j \neq i$ . Accept  $\mathbf{z}$  as  $\mathbf{x}^{(t)}$  with probability  $\min\{1, f_\phi(\mathbf{z})/f_\phi(\mathbf{x}^{(t-1)})\}$ , otherwise let  $\mathbf{x}^{(t)} = \mathbf{x}^{(t-1)}$ . The resulting sequence of point patterns is a realization of a Markov chain with invariant distribution  $f_\phi$ .

Although the estimation methods discussed below are applicable in a much more general setting (see Geyer 1998), it is sufficient for our purposes (and notationally somewhat simpler) to consider exponential families  $\{f_\theta: \theta \in \Theta\}$  of Gibbs point processes having the energy function

$$U_\theta(\mathbf{x}) = -\theta \cdot \mathbf{y}(\mathbf{x}), \tag{2.5}$$

where  $\mathbf{y}: \Omega \rightarrow \mathbb{R}^p$  is the canonical statistic, and  $\Theta = \{\theta \in \mathbb{R}^p : Z(\theta) < \infty\}$ . The log-likelihood of an observation  $\mathbf{x}$  is then

$$l(\theta) = \theta \cdot \mathbf{y}(\mathbf{x}) - \log Z(\theta). \tag{2.6}$$

### 2.2.1. Monte Carlo root-finding algorithms

In Penttinen (1984) the Newton–Raphson algorithm was considered for the maximization of the log-likelihood  $l$ . The values of the gradient  $\mathbf{g}(\theta) = \nabla l(\theta)$  and the Hessian  $\mathbf{H}(\theta) = \nabla^T \mathbf{g}(\theta)$  are required at the current parameter value  $\theta^{(k)}$  to perform the update

$$\theta^{(k+1)} = \theta^{(k)} - \mathbf{H}(\theta^{(k)})^{-1} \mathbf{g}(\theta^{(k)}). \tag{2.7}$$

By the usual formulae for exponential families they can be expressed as

$$\mathbf{g}(\theta) = \mathbf{y}(\mathbf{x}) - E_\theta \mathbf{y} \tag{2.8}$$

and

$$\mathbf{H}(\theta) = -\text{var}_\theta \mathbf{y}. \tag{2.9}$$

The *Monte Carlo Newton–Raphson* algorithm is obtained by replacing the expectation and the variance by their MCMC approximations. In each update an MCMC sample  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(T)}$  with target distribution  $f_{\theta^{(k)}}$  is drawn, the sample mean vector  $\bar{\mathbf{y}}$  and the sample covariance matrix  $\mathbf{S}_y$  of  $\mathbf{y}(\mathbf{x}^{(1)}), \dots, \mathbf{y}(\mathbf{x}^{(T)})$  are computed,  $\mathbf{g}(\theta^{(k)})$  in equation (2.7) is replaced by

$$\hat{\mathbf{g}}(\theta^{(k)}) = \mathbf{y}(\mathbf{x}) - \bar{\mathbf{y}} \tag{2.10}$$

and  $\mathbf{H}(\theta^{(k)})$  by

$$\hat{\mathbf{H}}(\theta^{(k)}) = -\mathbf{S}_y. \tag{2.11}$$

The MCMC approximations of  $\mathbf{g}$  and  $\mathbf{H}$  can naturally be used in any numerical root-finding algorithm. We have found the *Marquardt algorithm* (Marquardt 1963) to be particularly useful, since  $\mathbf{H}^{-1}$  can be extremely unstable at early stages of the iteration, especially when we have a high-dimensional parameter space and the initial value of  $\theta$  is far from the MLE. This is a well-known problem even when the ordinary (deterministic) Newton–Raphson method can be used, but it is made still more severe by the randomness of the MCMC approximations. In the Marquardt algorithm the Hessian is stabilized by

multiplying its diagonal by  $1 + \lambda$ , where  $\lambda$  is positive and tends to zero as the root is approached. The version given in Press *et al.* (1992, pp. 681–688) runs as follows. Define the merit function  $\chi^2$  by

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^p g_i(\boldsymbol{\theta})^2, \tag{3.8}$$

whereby the task of finding the root of  $\mathbf{g}$  is equivalent to minimizing  $\chi^2$ . Let  $\mathbf{I}_\lambda(\boldsymbol{\theta})$ ,  $\lambda \geq 0$ , be the inverse of the matrix obtained by multiplying the diagonal elements of  $\mathbf{H}(\boldsymbol{\theta})$  by  $1 + \lambda$ . Choose an initial parameter value  $\boldsymbol{\theta}^{(0)}$ , set  $\lambda = 0.001$ , and iterate the following steps for  $k = 0, 1, \dots$  until  $\chi^2$  decreases only by a negligible amount:

1. Given the current iterate  $\boldsymbol{\theta}^{(k)}$ , propose  $\boldsymbol{\theta}' = \boldsymbol{\theta}^{(k)} - \mathbf{I}_\lambda(\boldsymbol{\theta}^{(k)})\mathbf{g}(\boldsymbol{\theta}^{(k)})$  to be the next one.
2. If  $\chi^2(\boldsymbol{\theta}') \geq \chi^2(\boldsymbol{\theta}^{(k)})$ , refuse the proposal, increase  $\lambda$  by a factor of 10 and return to step 1.
3. Otherwise, accept  $\boldsymbol{\theta}'$  as the new state  $\boldsymbol{\theta}^{(k+1)}$  and decrease  $\lambda$  by a factor of 10.

Again, the *Monte Carlo Marquardt* algorithm is defined to be the version where the values of  $\mathbf{g}$  and  $\mathbf{H}$  are replaced by their MCMC approximations (2.10) and (2.11).

When considering the stopping rules one should note a fundamental difference between numerical root-finding algorithms and their Monte Carlo counterparts. Instead of deterministic sequences, these Monte Carlo algorithms produce Markov chains in the parameter space, and their convergence must be understood in the sense of weak convergence to a stationary distribution. The transition laws of these chains are highly complicated, depending on the behaviour of the MCMC sampler, and analytical convergence results are very difficult to obtain. However, the consistency of the MCMC method implies that, with large samples, the Monte Carlo algorithm approximates the behaviour of the corresponding deterministic algorithm.

### 2.2.2. Monte Carlo likelihood

Geyer and Thompson (1992) used importance sampling to obtain an MCMC approximation of the partition function over a range of  $\boldsymbol{\theta}$ s using a sample from just one distribution  $f_{\boldsymbol{\theta}_0}$ . This is based on the relation

$$\frac{Z(\boldsymbol{\theta})}{Z(\boldsymbol{\theta}_0)} = E_{\boldsymbol{\theta}_0}[\exp\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \cdot \mathbf{y}\}], \tag{2.13}$$

yielding

$$l(\boldsymbol{\theta}) - l(\boldsymbol{\theta}_0) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \cdot \mathbf{y}(\mathbf{x}) - \log E_{\boldsymbol{\theta}_0}[\exp\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \cdot \mathbf{y}\}]. \tag{2.14}$$

Thus, in principle, a sample from  $f_{\boldsymbol{\theta}_0}$  provides a Monte Carlo approximation of the entire log-likelihood, the *Monte Carlo likelihood*, and an approximate MLE can be obtained by maximizing the MCMC approximation of the right-hand side of (2.14). With a reasonable sample size, however, the approximation is accurate only in the vicinity of the pivot value  $\boldsymbol{\theta}_0$ , and an iteration of successive constrained optimizations is required. When  $\boldsymbol{\theta}_0$  is far from the

mode, the approximate log-likelihood may fail to have a maximum. Then also the constrained maximization in a high-dimensional space becomes complicated. For  $\theta_0$  near the mode, on the other hand, this method seems to surpass the Monte Carlo root-finding algorithms, which in turn provide simple and quick means of getting near the mode.

In order to evaluate the log-likelihood ratio  $l(\theta') - l(\theta)$  when  $\theta$  and  $\theta'$  are far apart, Geyer (1991) proposed a Monte Carlo likelihood approximation based on a sample from a mixture  $f = \sum_{m=1}^M f_{\theta_m}$  of distributions in the model family instead of just one. The sequence  $\theta_1, \dots, \theta_M$  of parameter values is chosen from the line segment between  $\theta$  and  $\theta_0$  so that the distributions of the canonical statistic have considerable overlap in the samples under any two consecutive values. The ratios  $Z(\theta_m)/Z(\theta_1)$ ,  $m = 2, \dots, M$ , needed to calculate the appropriate importance weights, are estimated by the method of ‘reverse logistic regression’ introduced in Geyer (1991).

### 2.2.3. Bayesian estimation

Let us now turn to Bayesian estimation, and assume that a prior distribution with density  $\pi$  is specified for  $\theta$ . Then the posterior density  $p(\theta|\mathbf{x})$  of  $\theta$  given the observation  $\mathbf{x}$  is proportional to  $\pi(\theta)f_{\theta}(\mathbf{x})$ , and the *posterior mode estimator* can be found by maximizing  $\log p(\theta|\mathbf{x})$ , which is equivalent to maximizing  $\log \pi(\theta) + l(\theta)$ . Hence the methods described in the previous subsections can be directly applied by defining

$$\mathbf{g}(\theta) = \nabla \log \pi(\theta) + \mathbf{y}(\mathbf{x}) - E_{\theta} \mathbf{y} \tag{2.15}$$

and

$$\mathbf{H}(\theta) = \nabla^T \nabla \log \pi(\theta) - \text{var}_{\theta} \mathbf{y} \tag{2.16}$$

for the root-finding algorithms (Section 2.2.1), or by adding  $\log \pi(\theta) - \log \pi(\theta_0)$  to the right-hand side of equation (2.14) for the Monte Carlo likelihood approximation (Section 2.2.2). It is only required that  $\pi$  and the first two differentials of  $\log \pi$  can be easily evaluated at any  $\theta$ .

In principle, the full posterior analysis could be conducted by MCMC sampling from the posterior distribution. In order to do this, however, the posterior ratios  $p(\theta'|\mathbf{x})/p(\theta|\mathbf{x})$  should be available for any current state  $\theta$  and for any possible update proposal  $\theta'$  in the sampler chain. Again, this is computationally equivalent to calculating the likelihood ratios, and in theory the Monte Carlo likelihood approximation based on a mixture (see the final paragraph of Section 2.2.2) could be applied. However, a large number of these update steps must be taken, and it is not practically possible to produce a new MCMC approximation at each. The only practical alternative seems to be that of having a mixture which covers that part of the parameter space where the posterior puts appreciable mass to begin with. In a low-dimensional case this is possible; an application was given in Higdon (1994, Chapter 7) in the context of image analysis. In a high-dimensional case we do not know of any practical way to choose a sufficient collection of parameter values for the mixture.

### 3. Pair potential estimation by Bayesian smoothing

Just as continuous functions can be approximated by step functions, so we can approximate an arbitrary pairwise interaction Gibbs process by a process of the *multi-scale family* (Penttinen 1984), in which pair potentials are step functions with given jump points  $0 = r_0 < r_1 < \dots < r_p$ . With the conditioning  $n(\mathbf{x}) = n$ , and using the parametrization

$$\phi = \sum_{i=1}^p -\theta_i \mathbf{1}_{(r_{i-1}, r_i]}, \tag{3.1}$$

we obtain an exponential family  $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta = (\mathbb{R} \cup \{-\infty\})^p\}$ , whose canonical statistic  $\mathbf{y} = (y_1, \dots, y_p)$  is given by the pair counts

$$y_i = y_i(\mathbf{x}) = \#\{(j, k) : 1 \leq j < k \leq n, r_{i-1} < \|x_j - x_k\| \leq r_i\}. \tag{3.2}$$

We can obtain an almost nonparametric estimate of the pair potential by fitting a multi-scale process of large order  $p$ . The problem is that with short intervals  $(r_{i-1}, r_i]$  only few observed point locations contribute to each component  $y_i$  of the sufficient statistic. Consequently, the sampling variance of  $y_i$  is large compared to its expectation, which in turn causes instability in the maximum likelihood estimation of  $\boldsymbol{\theta}$  (for an example, see Fig. 4).

Such a problem is familiar in image analysis, for example in the restoration of noisy pixel images. There a common remedy is to adopt a Bayesian approach and choose a prior distribution which gives higher probabilities to smoother images. This idea can be applied here as well. We specify a prior distribution  $\pi$  preferring such potentials  $\phi$  in which the jumps  $|\theta_i - \theta_{i-1}|$  are small. Analogously to the use of Markov random fields in image analysis, we choose  $\pi$  to be a Markov chain with zero mean jumps.

#### 3.1. The smoothing prior

First we need to decide on  $p$ , the resolution of the multi-scale potential, and on the locations  $r_i$  of the jump points. The radius of interaction  $r_p$  should be chosen so that the interaction beyond  $r_p$  is negligible. Often some background knowledge is available for the determination of  $r_p$ . Otherwise, a safe way would be to estimate the radius of interaction, by the method of van Lieshout and Baddeley (1996) for example, and then make  $r_p$  substantially larger than the estimated value. The choice of  $p$  determines how nonparametric the inference is. The computational load, of course, increases with  $p$ . A conventional choice for the other  $r_i$  ( $i = 1, \dots, p - 1$ ) is to have them equally spaced between  $r_0 = 0$  and  $r_p$ .

Second, the Markov chain prior  $\pi(\boldsymbol{\theta})$  needs to be specified. Expecting no interaction beyond  $r_p$ , it is convenient to start the chain from  $\theta_p$  with  $E_{\pi}\theta_p = 0$ . A simple choice is to take the ‘initial’ distribution and the transition kernels to be Gaussian with  $E_{\pi}(\theta_i | \theta_{i+1}) = \theta_{i+1}$ . The degree of smoothing is then determined by the variances

$$\sigma_i^2 = \begin{cases} \text{var}_{\pi}\theta_p, & \text{if } i = p, \\ \text{var}_{\pi}(\theta_i | \theta_{i+1}), & \text{otherwise.} \end{cases} \tag{3.3}$$

In general it is not reasonable to assume the  $\sigma_i^2$  to be equal. An extreme example, the hard-core process (e.g., Ripley 1977), corresponds to  $\theta_i = -\infty$  for  $r_i < R$ . The interaction parameters  $\eta_i = \exp(\theta_i)$ , on the other hand, are bounded between 0 and 1 for repulsive potentials, and typically the interaction function  $h = \exp(-\phi)$  behaves more regularly than the pair potential. Hence, a feasible simplification is to assume the conditional variances  $\text{var}_\pi(\eta_i|\eta_{i+1})$  equal rather than the  $\sigma_i^2$ . The induced expressions of the  $\sigma_i^2$  are available: if  $\theta \sim N(\mu_i, \sigma_i^2)$ , then  $\eta = \exp(\theta)$  follows the log-normal distribution (see Antle 1985) with variance  $\sigma^2 = \exp(2\mu_i + \sigma_i^2)\{\exp(\sigma_i^2) - 1\}$ , yielding

$$\sigma_i^2 = \log\left[\frac{1}{2} + \left\{\sigma^2 \exp(-2\mu_i) + \frac{1}{4}\right\}^{1/2}\right] =: s_\sigma(\mu_i). \tag{3.4}$$

We thus suggest a prior distribution defined by

$$\begin{aligned} \theta_p &\sim N(0, \sigma_p^2), \\ \theta_i|\theta_{i+1}, \dots, \theta_p &\sim N(\theta_{i+1}, \sigma_i^2), \end{aligned} \tag{3.5}$$

where  $\sigma_i^2 = s_\sigma(\theta_{i+1})$ , for  $i = 1, \dots, p - 1$ ,  $\sigma_p^2 = s_\sigma(0)$ , and the smoothing parameter  $\sigma$  determines how much weight we give to the prior compared to the data. The formulae for the joint prior distribution  $\pi(\boldsymbol{\theta})$  become somewhat complicated but still tractable. They are given in the Appendix.

In our prior the marginal variances  $\text{var}_\pi(\theta_i) = \sigma_i^2 + \dots + \sigma_p^2$  are larger for smaller  $i$ . This reflects our assumption that  $\theta_p$  should be nearly 0, whereas only vague prior knowledge is assumed of  $\theta_1$ . An alternative smoothing prior can be found in Arjas and Heikkinen (1997), where the two-sided conditionals  $p(\theta_i|\theta_{i-1}, \theta_{i+1})$  are specified.

For pair potentials that exhibit abrupt changes (such as the hard-core process) there is naturally the danger of oversmoothing. A more robust prior could be obtained by using the  $t$ -distributions; see Besag and Higdon (1993, Section 3), Besag *et al.* (1995, Section 5.6) or Besag and Higdon (1997, Section 2.4).

### 3.2. An algorithm for the posterior mode estimation

Next we describe a procedure for locating the posterior mode. First, it should be noticed that the unimodality of the posterior is difficult to check. For example, with a small prior variance  $\sigma^2$  one local mode could well be at  $\theta_i = 0$  for all  $i$ . The main interest lies, however, in the local mode, which is in best agreement with the data. Hence, we use the MLE (with a small modification) as the starting point of the iteration. Accepting  $\theta_i = -\infty$  for  $i \in I_0 = I_0(\mathbf{x}) := \{i : y_i(\mathbf{x}) = 0\}$ , the MLE always exists and is unique. Second, based on the considerations in Section 2.2 and on our experiences in modelling the data of Section 4, we propose that the Monte Carlo Marquardt algorithm is an efficient way of finding a pivot parameter for the Monte Carlo likelihood approximation sufficiently near the posterior mode. An outline of the suggested procedure is as follows:

1. Find an approximation  $\hat{\boldsymbol{\theta}}^{(ML)}$  to the MLE using the Monte Carlo Marquardt algorithm.
2. Let  $\boldsymbol{\theta}^{(0)}$  be a slight modification of  $\hat{\boldsymbol{\theta}}^{(ML)}$  such that  $\theta_i^{(0)} > -\infty$  for all  $i$ .

3. Find an approximation  $\hat{\boldsymbol{\theta}}^{(\text{MCM})}$  of the posterior mode using the Monte Carlo Marquardt algorithm initialized by  $\boldsymbol{\theta}^{(0)}$ .
4. Obtain the final estimate  $\hat{\boldsymbol{\theta}}$  using the Monte Carlo likelihood approximation with pivot parameter  $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}^{(\text{MCM})}$ .

Useful starting values for step 1 can be obtained by applying the Poisson approximation (Penttinen 1984, p. 40), yielding

$$\theta_i = \log \frac{y_i(\mathbf{x})}{K_i}, \tag{3.6}$$

where

$$K_i = \frac{\pi}{2\nu(E)} n^2 (r_i^2 - r_{i-1}^2). \tag{3.7}$$

The MLEs for  $\theta_i$ ,  $i \in I_0$ , are equal to  $-\infty$ . Hence, in step 1 we fix  $\theta_i = -\infty$  for  $i \in I_0$ , and update only the rest of the parameter values. In the smoothed estimate we require finite values for each  $\theta_i$ , and therefore we need to do the modification of step 2. A natural way is to transform into the interaction parameters  $\hat{\boldsymbol{\eta}}^{(\text{ML})} = \exp(\hat{\boldsymbol{\theta}}^{(\text{ML})})$ , whereby  $\hat{\eta}^{(\text{ML})} = 0$  for  $i \in I_0$ , and to construct  $\boldsymbol{\eta}^{(0)}$  as follows:

- If  $p \in I_0$ , let  $\eta_p^{(0)} = \frac{1}{2}$ , which is a compromise between  $\exp E_{\pi} \theta^{(0)} = 1$  and  $\eta^{(\text{ML})} = 0$ .
- Iteratively for  $i = p - 1, \dots, 1$ , if  $i \in I_0$  then let  $\eta_i^{(0)} = \eta_{i+1}^{(0)}/2$ , with the same reasoning as above.

Step 2 is completed by setting  $\boldsymbol{\theta}^{(0)} = \log \boldsymbol{\eta}^{(0)}$ .

Given the Monte Carlo approximation of the (log) posterior, step 4 consists of a routine numerical optimization, for which standard stopping rules can be applied. In steps 1 and 3, on the other hand, the randomness must be acknowledged when stating a stopping rule (see Section 2.2.1). The merit function  $\chi^2$  does not converge to zero, but ends up oscillating around a strictly positive value instead.

To get a rough idea of the expected behaviour of  $\chi^2$  near the mode, let us first note that the central limit theorem holds for the MCMC approximations of the  $g_i(\boldsymbol{\theta})$ . Given  $\boldsymbol{\theta} \in \Theta$ , let  $I_{\infty} = \{i : \theta_i = -\infty\}$ . Then the set  $\Omega_n^+ = \{\mathbf{x} \in \Omega_n : f_{\boldsymbol{\theta}}(\mathbf{x}) > 0\}$  consists of those patterns  $\mathbf{x}$  for which  $I_0(\mathbf{x}) = I_{\infty}$ . Within this set the density  $f_{\boldsymbol{\theta}}$  is bounded from above by  $\exp(\theta_{\max, y_{\max}})/Z(\boldsymbol{\theta})$  and from below by  $\exp(\theta_{\min, y_{\max}})/Z(\boldsymbol{\theta})$ , where  $y_{\max} = n(n + 1)/2$  is the maximal pair count,  $\theta_{\min} = \min_{i \notin I_{\infty}} \theta_i$  and  $\theta_{\max} = \max\{\theta_1, \dots, \theta_p, 0\}$ . Hence the conditions of Corollary 3 in Tierney (1994) are satisfied, the sampler chain is uniformly ergodic, and the central limit theorem holds for all square-integrable functionals of  $\mathbf{x}$  (Tierney 1994, Theorem 5).

In particular, the MCMC approximation  $\hat{g}_i(\boldsymbol{\theta})$  based on a sample of size  $T$  can be regarded as an approximate realization from the normal distribution with mean  $g_i(\boldsymbol{\theta})$  and variance  $\tau_i^2/T$ , where  $\tau_i^2 = \tau_i^2(\boldsymbol{\theta})$  is unknown but finite. If  $\boldsymbol{\theta}$  were the root of  $\mathbf{g}$  and the  $\hat{g}_i(\boldsymbol{\theta})$ ,  $i = 1, \dots, p$ , were independent, then the MCMC approximation

$$\hat{\chi}^2(\boldsymbol{\theta}) = \sum_{i=1}^p \frac{\hat{g}_i(\boldsymbol{\theta})^2}{\tau_i^2} \tag{3.8}$$

of the normalized merit function would follow the central  $\chi_p^2$  distribution.

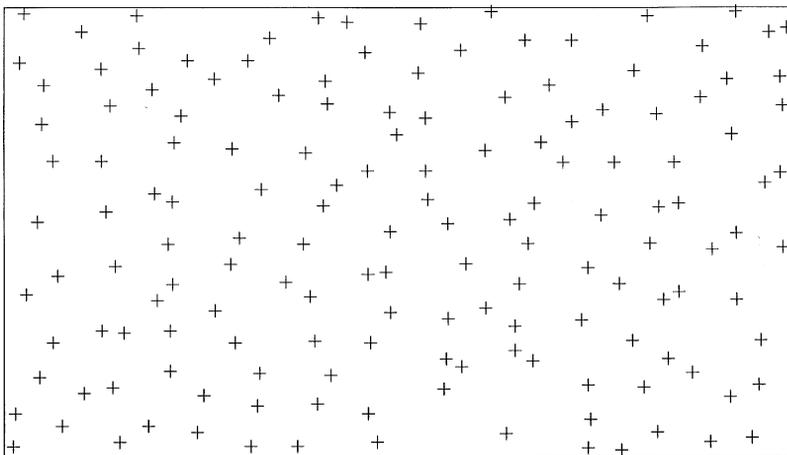
Based on these considerations, we suggest the following stopping rule for steps 1 and 3. Estimate (by the method of Geyer 1992, for example) the Monte Carlo variances  $\tau_i^2$  of the pair counts  $y_i$  from an MCMC sample under the initial  $\theta$ -value, and substitute these estimates for the  $\tau_i^2$  in equation (3.8). Stop when  $\hat{\chi}^2$  reaches the 95th percentile of the  $\chi_p^2$  distribution. This rule is based on very crude approximations: The  $\hat{g}_i$  clearly cannot be independent, and we cannot expect the  $\tau_i^2$  be the same at the root of  $\mathbf{g}$  as at the initial value. Also the 95th percentile seems quite liberal. But it is not necessary to be very accurate at this stage, because the aim is only to get reasonably near the root. It actually turned out that even this stopping rule was unnecessarily strict in our example (see Section 4).

Finally, we should mention that the estimate  $\hat{\theta}$  can always be checked by comparing the absolute values of the obtained  $\hat{g}_i(\hat{\theta})$  to their estimated Monte Carlo errors.

## 4. Example

We applied our procedure to the rabbit's eye amacrine cells data presented in Diggle and Gratton (1984). The study area  $E$  is a rectangle of  $1070 \times 600 \mu\text{m}$ ; the locations of the 152 cells within  $E$  are shown in Fig. 1.

We used the multi-scale process of order  $p = 30$ , with change-points equally spaced between  $r_0 = 0$  and  $r_p = 120 \mu\text{m}$ . In each simulation we used the Metropolis algorithm with 110 000 basic update steps; after a burn-in period of 10 000 steps every 100th realization was collected to form a sample of size  $T = 1000$ . Periodic boundaries were used both in calculation of the  $y_i$ s and in simulation. Here we report briefly steps 3 and 4 of the run with  $\sigma^2 = 0.01$ .



**Figure 1.** Locations of 152 amacrine cells in a rabbit's eye; from Diggle and Gratton (1984).

For step 3 of our algorithm the  $\chi^2$  approximation suggested in Section 3.2 yields stopping rule  $\hat{\chi}^2 \leq 43.77$  (the 95th percentile of the  $\chi^2_{30}$  distribution). This was reached in seven iterations of the Monte Carlo Marquardt algorithm with 13 evaluations of  $\hat{\chi}^2$ . The values are as follows with those corresponding to the accepted  $\theta$ -values underlined (see steps 2 and 3 of the Marquardt algorithm):

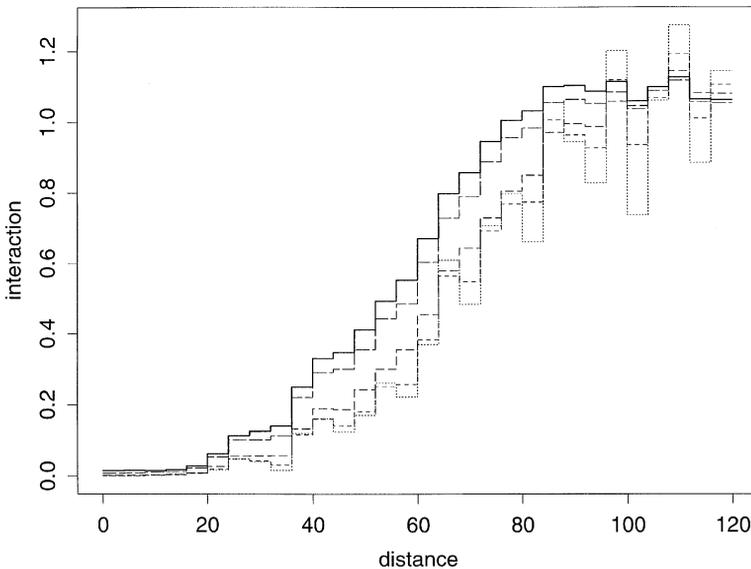
93 918, 290 429, 15 005, 484, 168, 119, 63, 59, 116, 139, 101, 65, 27.

The  $\eta$ -values of some selected iterations are shown in Fig. 2. The final estimate, using the Monte Carlo likelihood approximation, is practically equal to that given by the Monte Carlo Marquardt algorithm.

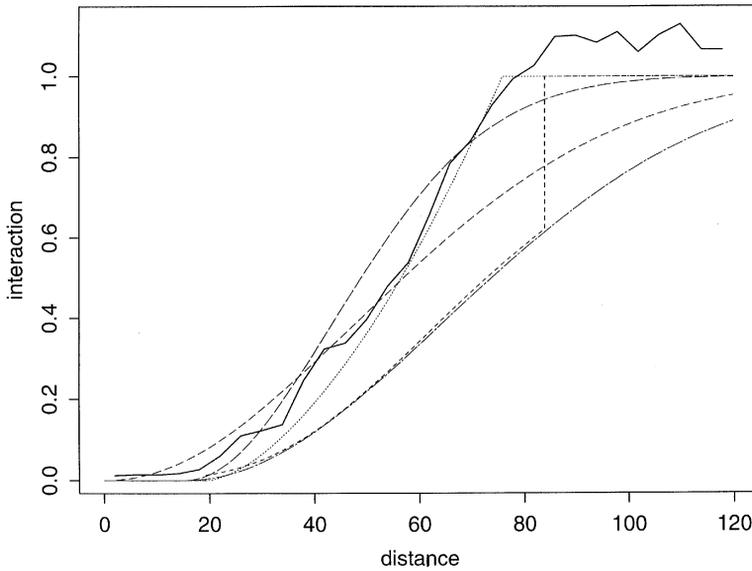
In Fig. 3 our estimate is compared to earlier (parametric) interaction function estimates, based on a summary in Särkkä (1993). For easier comparison we have constructed a continuous version of our estimate by joining the midpoints of the steps. For intermediate distances our estimate seems to agree best with that of Diggle and Gratton. There is some evidence of attraction at longer distances, which may reflect higher-order interactions.

We also estimated the posterior modes corresponding to milder smoothings  $\sigma^2 = 0.05$ ,  $\sigma^2 = 0.2$  and  $\sigma^2 = 1$ . The estimates are compared in Fig. 4. Clearly, the posterior modes are far from being simple smoothings of the MLE.

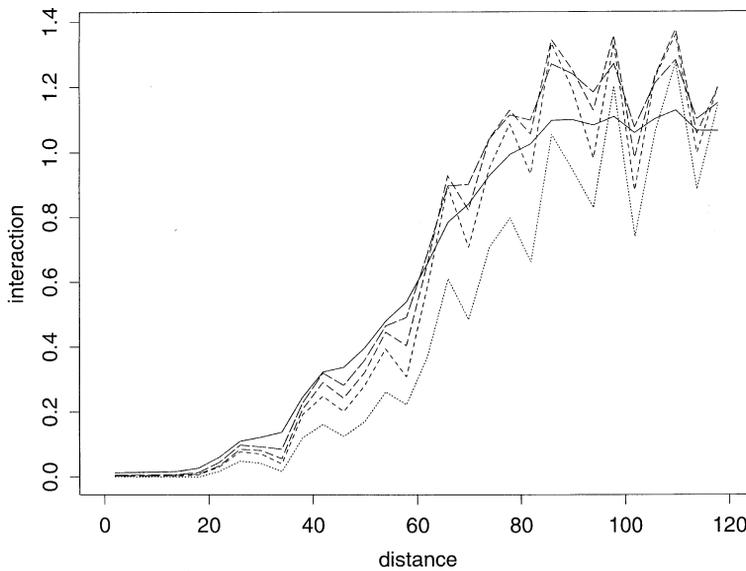
Finally, we performed a related simulation study. Nine realizations were drawn using Diggle and Gratton's potential, and the above procedure (with  $p = 20$  and  $\sigma^2 = 0.01$ ) was



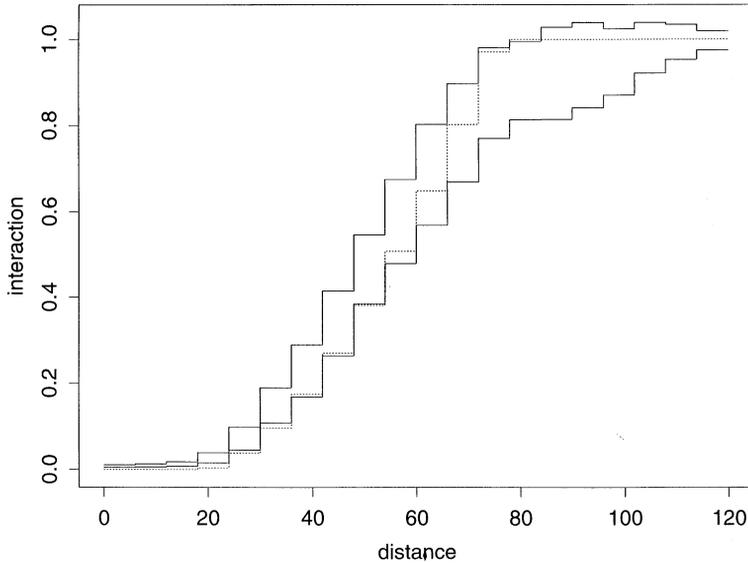
**Figure 2.** The  $\eta$  values after selected iterations of the Monte Carlo Marquardt algorithm for the posterior mode estimate with  $\sigma^2 = 0.01$ : initial values  $\cdots\cdots$ , first iteration  $----$ , second iteration  $-\cdot-\cdot-$ , third iteration  $----$ , and final values  $—$ .



**Figure 3.** Continuous version (see text) of our posterior mode estimate (solid line) compared to various parametric estimates: Diggle and Gratton (1984)  $\cdots\cdots$ , Fiksel and Stoyan (unpublished)  $-\cdot-\cdot-$ , Ogata and Tanemura (1984)  $-\cdot-\cdot-$ , Penttinen (unpublished)  $-\cdot-\cdot-$ , and Särkkä (1993)  $-\cdot-\cdot-$ .



**Figure 4.** Continuous versions of the posterior mode estimates under various values of the smoothing parameter:  $\sigma^2 = 0.01$   $-\cdot-\cdot-$ ,  $\sigma^2 = 0.05$   $-\cdot-\cdot-$ ,  $\sigma^2 = 0.2$   $-\cdot-\cdot-$ ,  $\sigma^2 = 1$   $-\cdot-\cdot-$ , and  $\sigma^2 = \infty$  (MLE)  $\cdots\cdots$ .



**Figure 5.** A step function approximation ( $p = 20$ ) of Diggle and Gratton's interaction function (broken line), and stepwise means plus and minus standard deviations of the posterior mode estimates under  $\sigma^2 = 0.01$  from nine patterns simulated from Diggle and Gratton's model (solid lines).

applied to each. The stepwise means plus and minus standard deviations of the estimates are shown in Fig. 5.

## 5. Discussion

We have presented an approach to the nonparametric estimation of the pair potential function. As such it is comparable to that of Diggle *et al.* (1987). In their approach the nonparametric estimate of the pair correlation function and an approximate equation are used, and the properties of the final estimator are not directly accessible. Our method is more direct and, in principle, allows for full posterior analysis. Furthermore, unlike analytic approximations, the MCMC approximations can be made arbitrarily accurate by increasing the sample size. On the other hand, the computational effort required in our approach is substantial, even for the posterior mode.

It may be argued that a step function is not a proper estimator of the pair potential, which is usually assumed continuous. The major argument in favour of step functions is in their simplicity; among other things, they allow for the simple form of the sufficient statistic  $\mathbf{y}$ . Specification of the prior can also be done in a natural way. Rather than using splines, for example, we see variable change-points as a natural extension towards more

flexibility. The posterior mean would then usually be a smooth curve (see Arjas and Gasbarra 1994).

We conditioned on the number of points only in order to clarify the basic idea. In the unconditional case an additional parameter (the chemical activity) would have to be estimated, and some constraints would be required for the parameters in order to stay in the class of integrable densities.

## Acknowledgements

The idea of Bayesian smoothing arose in a seminar with Elja Arjas and his research group. The application of the Marquardt algorithm instead of the Newton–Raphson method was suggested by Charles Geyer. Fruitful discussions with Jesper Møller on point processes have improved the paper, as have the helpful comments of the Associate Editor and a referee. During the study JH had an opportunity to work under the supervision of Peter Green at the Department of Mathematics, University of Bristol. Financially, this work has been supported by the Academy of Finland (grants 1031177 and 1051328) and by the University of Jyväskylä.

## Appendix: Differentials of $\log \pi$

Below we shall derive some formulae concerning the joint distribution  $\pi(\boldsymbol{\theta})$  defined by (3.5). Let

$$A(\theta) = \sigma^2 \exp(-2\theta),$$

$$s(\theta) = \log\left[\frac{1}{2} + \{A(\theta) + \frac{1}{4}\}^{1/2}\right],$$

and let  $s_0 = s(0)$ . Then the joint density is defined by

$$\pi(\boldsymbol{\theta}) \propto \left[ \prod_{i=2}^p s(\theta_i)^{-1/2} \exp\left\{-\frac{1}{2}s(\theta_i)^{-1}(\theta_{i-1} - \theta_i)^2\right\} \right] \exp\left\{-\frac{1}{2}s_0^{-1}\theta_p^2\right\}$$

and its logarithm is

$$\log \pi(\boldsymbol{\theta}) = c - \frac{1}{2} \left[ \sum_{i=2}^p \{\log s(\theta_i) + s(\theta_i)^{-1}(\theta_{i-1} - \theta_i)^2\} + s_0^{-1}\theta_p^2 \right],$$

where  $c$  does not depend on  $\boldsymbol{\theta}$ . The gradient of the log-prior is given by

$$\frac{\partial}{\partial \theta_1} \log \pi(\boldsymbol{\theta}) = \frac{\theta_2 - \theta_1}{s(\theta_2)},$$

$$\frac{\partial}{\partial \theta_i} \log \pi(\boldsymbol{\theta}) = \frac{\theta_{i-1} - \theta_i - \frac{1}{2}s'(\theta_i)}{s(\theta_i)} + \frac{\theta_{i+1} - \theta_i}{s(\theta_{i+1})} + \frac{s'(\theta_i)}{2} \left\{ \frac{\theta_{i-1} - \theta_i}{s(\theta_i)} \right\}^2$$

for  $i = 2, \dots, p - 1$ , and

$$\frac{\partial}{\partial \theta_p} \log \pi(\boldsymbol{\theta}) = \frac{\theta_{p-1} - \theta_p - \frac{1}{2}s'(\theta_p)}{s(\theta_p)} - \frac{\theta_p}{s_0} + \frac{s'(\theta_p)}{2} \left\{ \frac{\theta_{p-1} - \theta_p}{s(\theta_p)} \right\}^2,$$

where

$$s'(\theta) = \{4A(\theta) + 1\}^{-1/2} - 1.$$

The Hessian is a (symmetric) tridiagonal matrix with diagonal elements

$$\frac{\partial^2}{\partial \theta_1^2} \log \pi(\boldsymbol{\theta}) = -\frac{1}{s(\theta_2)},$$

$$\begin{aligned} \frac{\partial^2}{\partial \theta_i^2} \log \pi(\boldsymbol{\theta}) = & -\frac{1}{s(\theta_{i+1})} - \frac{1}{s(\theta_i)} - \frac{s''(\theta_i)}{2s(\theta_i)} + \frac{1}{2} \left\{ \frac{s'(\theta_i)}{s(\theta_i)} \right\}^2 \\ & + \frac{s''(\theta_i)}{2} \left\{ \frac{\theta_{i-1} - \theta_i}{s(\theta_i)} \right\}^2 - \frac{2s'(\theta_i)(\theta_{i-1} - \theta_i)}{s(\theta_i)^2} - \frac{\{s'(\theta_i)(\theta_{i-1} - \theta_i)\}^2}{s(\theta_i)^3} \end{aligned}$$

for  $i = 2, \dots, p - 1$ , and

$$\begin{aligned} \frac{\partial^2}{\partial \theta_p^2} \log \pi(\boldsymbol{\theta}) = & -\frac{1}{s_0} - \frac{1}{s(\theta_p)} - \frac{s''(\theta_p)}{2s(\theta_p)} + \frac{1}{2} \left\{ \frac{s'(\theta_p)}{s(\theta_p)} \right\}^2 \\ & + \frac{s''(\theta_p)}{2} \left\{ \frac{\theta_{p-1} - \theta_p}{s(\theta_p)} \right\}^2 - \frac{2s'(\theta_p)(\theta_{p-1} - \theta_p)}{s(\theta_p)^2} - \frac{\{s'(\theta_p)(\theta_{p-1} - \theta_p)\}^2}{s(\theta_p)^3}, \end{aligned}$$

and (non-zero) sub-diagonal elements

$$\frac{\partial^2}{\partial \theta_i \partial \theta_{i-1}} \log \pi(\boldsymbol{\theta}) = \frac{1}{s(\theta_i)} + \frac{s'(\theta_i)(\theta_{i-1} - \theta_i)}{s(\theta_i)^2},$$

for  $i = 2, \dots, p$ , where

$$s''(\theta) = 4A(\theta)\{4A(\theta) + 1\}^{-3/2}.$$

## References

- Antle, C.E. (1985) Lognormal distribution. In S. Kotz, N.L. Johnson and C.B. Read (eds), *Encyclopedia of Statistical Sciences*, vol. 5, pp. 134–136. New York: Wiley.
- Arjas, E. and Gasbarra, D. (1994) Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statist. Sinica*, **4**, 505–524.
- Arjas, E. and Heikkinen, J. (1997) An algorithm for nonparametric Bayesian estimation of a Poisson intensity. *Comput. Statist.*, **12**, 385–402.
- Baddeley, A.J. and Møller, J. (1989) Nearest-neighbour Markov point processes and random sets. *Internat. Statist. Rev.*, **57**, 89–121.

- Baddeley, A.J. and van Lieshout, M.N.M. (1995) Area-interaction point processes. *Ann. Inst. Statist. Math.*, **47**, 601–619.
- Baddeley, A. J., van Lieshout, M.N.M. and Møller, J. (1996) Markov properties of cluster processes. *Adv. Appl. Probab.*, **28**, 346–355.
- Besag, J. (1986) On the statistical analysis of dirty pictures (with discussion). *J. Roy. Statist. Soc. Ser. B*, **48**, 259–302.
- Besag, J. and Higdon, D. (1993) Bayesian inference for agricultural field experiments. *Bull. Internat. Statist. Inst.*, **55**, 121–136.
- Besag, J. and Higdon, D. (1997) Bayesian inference for agricultural field experiments. Discussion Paper 97/17. Institute of Statistics and Decision Sciences, Duke University.
- Besag, J., Green, P.J., Higdon, D. and Mengersen, K. (1995) Bayesian computation and stochastic systems (with discussion). *Statist. Sci.*, **10**, 3–66.
- Carter, D.S. and Prenter, P.M. (1972) Exponential spaces and counting processes. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **21**, 1–19.
- Denison, D.G.T., Mallick, B.K. and Smith, A.F.M. (1998) Automatic Bayesian curve fitting. *J. Roy. Statist. Soc. Ser. B*, **60**, 333–350.
- Diggle, P.J. and Gratton, R.J. (1984) Monte Carlo methods of inference for implicit statistical models (with discussion). *J. Roy. Statist. Soc. Ser. B*, **46**, 193–227.
- Diggle, P.J., Gates, D.J. and Stibbard, A. (1987) A nonparametric estimator for pairwise-interaction point processes. *Biometrika*, **74**, 763–770.
- Diggle, P.J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. and Tanemura, M. (1994) On parameter estimation for pairwise interaction point processes. *Internat. Statist. Rev.*, **62**, 99–117.
- Fiksel, T. (1988) Estimation of interaction potentials of Gibbsian point processes. *Statistics*, **19**, 77–86.
- Gates, D.J. and Westcott, M. (1986) Clustering estimates for spatial point distributions with unstable potentials. *Ann. Inst. Statist. Math.*, **38**, 123–135.
- Geyer, C.J. (1991) Estimating normalizing constants and reweighting Markov chain Monte Carlo. Technical Report 568, School of Statistics, University of Minnesota.
- Geyer, C.J. (1992) Practical Markov chain Monte Carlo (with discussion). *Statist. Sci.*, **7**, 473–511.
- Geyer, C.J. (1998) Likelihood inference for spatial point processes. In O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout (eds), *Stochastic Geometry, Likelihood and Computation*, Monogr. Statist. Appl. Probab. 80, pp. 79–140. Boca Raton, FL: Chapman & Hall/CRC.
- Geyer, C.J. and Møller, J. (1994) Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Statist.*, **21**, 359–373.
- Geyer, C.J. and Thompson, E.A. (1992) Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *J. Roy. Statist. Soc. Ser. B*, **54**, 657–699.
- Geyer, C.J. and Thompson, E.A. (1995) Annealing Markov chain Monte Carlo with applications to ancestral inference. *J. Amer. Statist. Assoc.*, **90**, 909–920.
- Green, P.J. (1995) Reversible jump MCMC and Bayesian model determination. *Biometrika*, **82**, 711–732.
- Heikkinen, J. and Arjas, E. (1998) Non-parametric Bayesian estimation of a spatial Poisson intensity. *Scand. J. Statist.*, **25**, 435–450.
- Heikkinen, J. and Arjas, E. (1999) Modeling a Poisson forest in variable elevations: a nonparametric Bayesian approach. *Biometrics*, **55**, 738–745.
- Higdon, D. (1994) Spatial applications of Markov chain Monte Carlo for Bayesian inference. Ph.D. thesis, Department of Statistics, University of Washington.
- Kelly, F.P. and Ripley, B.D. (1976) A note on Strauss's model for clustering. *Biometrika*, **63**, 357–360.
- Marquardt, D.W. (1963) An algorithm for least squares estimation of non-linear parameters. *J. Soc.*

- Indust. Appl. Math.*, **11**, 431–441.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087–1092.
- Møller, J. (1998) Markov chain Monte Carlo and spatial point processes. In O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout (eds), *Stochastic Geometry, Likelihood and Computation*, Monogr. Statist. Appl. Probab. 80, pp. 141–172. Boca Raton, FL: Chapman & Hall/CRC.
- Ogata, Y. and Tanemura, M. (1984) Likelihood analysis of spatial point patterns. *J. Roy. Statist. Soc. Ser. B*, **46**, 496–518.
- Penttinen, A. (1984) *Modelling Interactions in Spatial Point Patterns: Parameter Estimation by the Maximum Likelihood Method*, doctoral thesis. Jyväskylä Stud. Comput. Sci. Econom. Statist. 7. Jyväskylä: Department of Statistics, University of Jyväskylä.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992) *Numerical Recipes in C*, 2nd edn. Cambridge: Cambridge University Press.
- Preston, C.J. (1977) Spatial birth-and-death processes. *Bull. Internat. Statist. Inst.*, **46**, 371–391.
- Ripley, B.D. (1977) Modelling spatial patterns (with discussion). *J. Roy. Statist. Soc. Ser. B*, **39**, 172–212.
- Ripley, B.D. (1979) Simulating spatial patterns: dependent samples from a multivariate density. *J. Roy. Statist. Soc. Ser. C*, **28**, 109–112.
- Ripley, B.D. (1988) *Statistical Inference for Spatial Processes*. Cambridge: Cambridge University Press.
- Särkkä, A. (1993) *Pseudo-likelihood Approach for Pair Potential Estimation of Gibbs Processes*, doctoral thesis. Jyväskylä Stud. Comput. Sci. Econom. Statist. 22. Jyväskylä: Department of Statistics, University of Jyväskylä.
- Stoyan, D., Kendall, W.S. and Mecke, J. (1995) *Stochastic Geometry and its Applications*, 2nd edn. New York: Wiley.
- Tierney, L. (1994) Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.*, **22**, 1701–1762.
- van Lieshout, M.N.M. and Baddeley, A.J. (1996) A nonparametric measure of spatial interaction in point patterns. *Statist. Neerlandica*, **50**, 344–361.

Received October 1995 and revised August 1997