Computational Statistics and Data Analysis 71 (2014) 530–541

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Bayesian analysis of a Gibbs hard-core point pattern model with varying repulsion range

T. Rajala [a,*], A. Penttinen [b]

[a] Helsinki Institute for Information Technology, Department of Information and Computer Science, P.O. Box 15400, 00076 Aalto, Finland

[b] Department of Mathematics and Statistics, P.O. Box 35, 40014 University of Jyväskylä, Finland

ARTICLE INFO

Article history:
Received 19 January 2012
Received in revised form 24 August 2012
Accepted 25 August 2012
Available online 5 September 2012

Keywords:
Hard-core point process
Inhomogeneous
Gaussian process regularisation
Bayesian analysis
Sand Martin's nests

ABSTRACT

A Bayesian solution is suggested for the modelling of spatial point patterns with inhomogeneous hard-core radius using Gaussian processes in the regularization. The key observation is that a straightforward use of the finite Gibbs hard-core process likelihood together with a log-Gaussian random field prior does not work without penalisation towards high local packing density. Instead, a nearest neighbour Gibbs process likelihood is used. This approach to hard-core inhomogeneity is an alternative to the transformation inhomogeneous hard-core modelling. The computations are based on recent Markovian approximation results for Gaussian fields. As an application, data on the nest locations of Sand Martin (*Riparia riparia*) colony[1] on a vertical sand bank are analysed.

1. Introduction

Observed spatial point patterns are usually assumed to be spatially stationary or, in the finite case, spatially homogeneous. However, data consisting of point locations may reveal inhomogeneity in point density, in local structure, or in both.

The present paper deals with Bayesian modelling of finite spatial point patterns with inhibition between the points. It is assumed that the range of inhibition is spatially inhomogeneous, and this inhomogeneity is modelled through a latent Gaussian process. As a result a posterior description of the background inhomogeneity is obtained as well as predictions for further points in the point pattern.

A motivating example consists of a point pattern of Sand Martin's (*Riparia riparia*) nests on a vertical sand bank, see Fig. 1. The figure reveals that the nest holes are highly packed but the packing density varies, partly according to the unobserved variation of sand composition in the bank. Further examples can be found e.g. in the fields of solid matter physics (Hahn et al., 2003) and physiology (Nielsen and Vedel Jensen, 2004).

The mainstream of the modelling of inhomogeneous point patterns is focused on the intensity estimation for inhomogeneous Poisson processes, see e.g. the discussion in Guttorp and Thorarinsdottir (2011). Because of the nature of the Poisson process, local structures such as inhibition between the points are excluded. Illian et al. (in press) show that interaction-like effect can be introduced to the Poisson process by adding spatially continuous morphological summaries as "constructed covariates" for the intensity. In the non-Poisson setting, Vedel Jensen and Nielsen (2000) consider a model where a homogeneous Markov point process is transformed to an inhomogeneous one. This process class is called

* Correspondence to: Department of Information and Computer Science, P.O. Box 15400, 00076 Aalto, Finland. Tel.: +358 400 695 103.
E-mail addresses: tuomas.rajala@iki.fi (T. Rajala), antti.k.penttinen@jyu.fi (A. Penttinen).

[1] Dataset is attached to the online version.

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.08.014

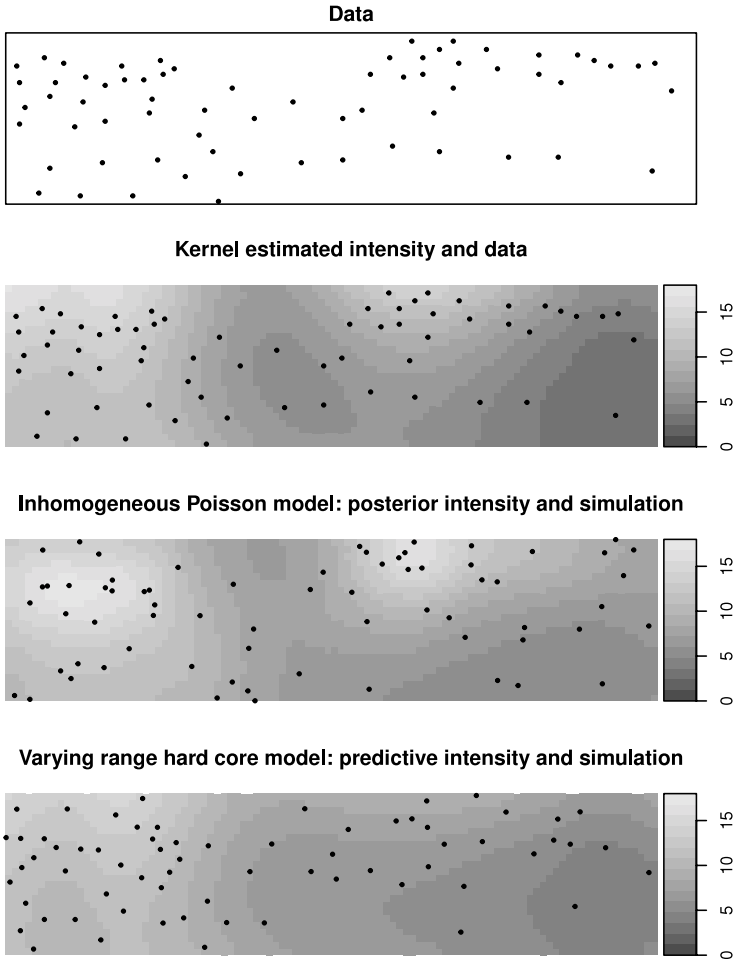

Fig. 1. From the top down: Locations of Sand Martin's nests on a sand bank as will be described in Section 4; data overlaid on a kernel smoothed intensity field; a simulation of an inhomogeneous Poisson process model overlaid on the mean posterior intensity field; simulation from the new varying range hard-core model overlaid on a mean posterior predictive intensity field.

"Transformation inhomogeneous Markov" (TIM), see also Nielsen and Vedel Jensen (2004) and Hahn et al. (2003). The use of transformation is a modification of the idea by Sampson and Guttorp (1992) for geostatistical random fields, preceded by the time acceleration models in survival analysis. The inference for the TIM model is based on finding the most likely transformation. As an alternative to TIM modelling, Berthelsen and Møller (2008) apply Markov point processes defined in terms of inhomogeneous self-potential and pairwise interaction function giving a Bayesian solution.

In the present paper it is assumed that the points in a point pattern have a hard-core with smoothly varying radius. The pattern can be thought to be the set of centre points of non-intersecting spherical objects of variable size. This model is called a *hard spheres* model, and the radius at a location is tied to the physical size of the object, see e.g. Månsson and Rudemo (2002). An alternative construction is to associate each point with a radius and define the hard-core area by excluding other points within the sphere. Then the variable radius can be addressed to the point location. This process is an extension of the Markov *hard-core* process (finite Gibbs hard-core process). We adopt the latter point of view.

The Markov hard-core process is defined by the self-potential and pairwise interaction, and both of these contribute to the Papangelou conditional intensity function, see Illian et al. (2008, p. 149) and van Lieshout (2000, p. 39). However, if we are close to the case of high packing, then the pairwise interaction (hard-core) dominates, and the spatial distribution of the hard-core radius has a primary effect. Our suggestion for modelling varying radius hard-core Gibbs processes is by means of latent smooth Gaussian processes, see Rasmussen and Williams (2006): the hard-core distance for each point is determined by a random function depending on this Gaussian process.

Two problems will be met in fitting the varying radius hard-core process. The first problem is the intractable scaling factor in the hard-core Markov point process likelihood which in this case depends also on the unknown hard-core radius function. The known solutions, importance sampling applied by Bognar (2005) and auxiliary variable method by Møller et al. (2006), lead to demanding computation. Also, the approximation of the likelihood with the pseudo-likelihood function (Jensen and Møller, 1991) depends on the unknown latent function. Secondly, although the synthesis (i.e. simulation of the

model) gives reasonable realisations, the analysis (estimation) does not work. The reason is that the Gibbs hard-core model contains locally just the upper limit information on the latent function. The problem is solved using the information on the local behaviour, the nearest neighbour distances.

Here a new inferential problem of general nature is presented, motivated by a data set on the nest colonies of the Sand Martins. The objective is to construct a point process model preserving the hard-core nature and allowing heterogeneity in the hard-core structure, to describe the spatial variation in the hard-core structure, and to carry out computation.

A solution is suggested assuming that the number of points in the study region is constant. This solution uses the local nearest neighbour distances which carry out information on the spatial variation of the hard-core radius. The smoothness of the hard-core radius field is inherited from a smoothly varying latent random function. Further, a tendency towards high packing is assumed.

This suggestion contains two parameters, one for smoothness and another one for the level of packing, which are difficult to determine from the point pattern simultaneously with the latent random field. The suggested two-stage procedure contains a "model selection" step which determines the smoothness and packing density, and posterior computation step for the hard-core random field.

In the present paper, Section 2 introduces the mathematical framework, defines the simple inhomogeneous hard-core model and illustrates the synthesis of the model. Section 3 describes the Bayesian analysis, solves the inversion problem related to the latent structure, and gives a simulation experiment. Bayesian analysis of the Sand Martin's nest data is conducted in Section 4, whilst Section 5 is for discussion.

2. Definition of the model and synthesis

2.1. Finite Gibbs model for hard-core point patterns

Assume that $\mathbf{x} = \{x_1, \ldots, x_n\}$ is a point pattern (point configuration) with the points $x_i \in D$ where $D \subset \mathbb{R}^2$ is a bounded window of observation and $n = 1, 2, \ldots$. Let $n_B(\mathbf{x}) := \#(\mathbf{x} \cap B)$ denote the number of points in a bounded set $B \subseteq \mathbb{R}^2$. A *point process* is a random variable that takes values in the configuration space such that $n_B(\mathbf{x}) < \infty$ for any set $B \subseteq D$. Assume further that the process is simple, i.e. multiple points are not possible.

A finite *Gibbs point process* with variable number of points has a parametric density

$$f(\mathbf{x}; \theta) = z_\theta^{-1} \exp(-U_\theta(\mathbf{x})) \tag{1}$$

with respect to the unit rate Poisson point process. The function U_θ, with some parameter vector θ, is called the total energy associated with the point pattern $\mathbf{x}$. The normalising constant z_θ is usually intractable. If the number of points n is fixed, then the conditional density is defined with respect to the binomial process in D with n points. For details of the construction of (finite) Gibbs processes, see Stoyan et al. (1995), Illian et al. (2008), van Lieshout (2000) or Møller and Waagepetersen (2004).

Denote the Euclidean distance with $d(\cdot, \cdot)$, and write $d_i := \min_{j \neq i} d(x_i, x_j)$ for the nearest neighbour distance of x_i in a pattern $\mathbf{x}$ with $n_D(\mathbf{x}) > 1$. A finite pairwise interaction model with $n_D(\mathbf{x}) = n$ points (n fixed), and with the parameter of interaction R, is defined through the total energy

$$U_R(\mathbf{x}) = \sum_{i<j} \phi_R(x_i, x_j),$$

where ϕ_R is a non-negative function modelling interactions between the point pairs. The finite *Gibbs hard-core model* is a pairwise interaction process for which the interaction function is

$$\phi_R(x_i, x_j) = \begin{cases} \infty, & \text{if } d(x_i, x_j) < R, \\ 0 & \text{otherwise.} \end{cases}$$

The model assigns zero density to a point pattern whenever the pattern has at least two points with interpoint distance less than R. This is equivalent to the event that the hard-core rule $\{\exists i : d_i < R\}$ is violated, leading to the presentation

$$U_R(\mathbf{x}) = \sum_{i=1}^{n} \tilde{\phi}_R(x_i, \mathbf{x}_{-i}) \tag{2}$$

with the interaction function

$$\tilde{\phi}_R(x_i, \mathbf{x}_{-i}) = \begin{cases} \infty, & \text{if } d_i < R, \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

with $\mathbf{x}_{-i} = \mathbf{x} \setminus \{x_i\}$.

2.2. Hierarchical Gibbs model with varying hard-core radius

Let us consider point configurations with varying, location-dependent hard-core radius, which is assumed to vary smoothly, see Fig. 1 as an example. The approach is hierarchical.

We base the varying range parameter on a latent *Gaussian random field* (GRF), a flexible family of models for dependent data in both spatial and higher dimensional problems (Cressie, 1993; Rasmussen and Williams, 2006; Gelfand et al., 2010). A GRF $\mathbf{Y} := \{s \in D : Y(s)\}$ is a random process that assigns a random value to each location s, and for which each finite vector $(Y(s_1), Y(s_2), \ldots, Y(s_N))$ has a joint Gaussian distribution for any $N \geq 1$. We use in the present paper a stationary and isotropic GRF model with constant mean μ and covariance between locations of distance t apart given by the Matérn covariance function

$$C(t) \propto \tau^{-1}(\kappa t)^{\nu} K_{\nu}(\kappa t).$$

The precision parameter τ is the inverse of the variance, ν is related to the differentiability of the field, K_{ν} is the modified Bessel function of the second kind, and κ determines the spatial scale. For interpretability, following Lindgren et al. (2011), the scale parameter κ is replaced with a range parameter $r = \sqrt{8\nu}/\kappa$ in order to obtain the correlation approximately of size 0.1 at a distance r for any value of parameter ν. Both ν and r affect the smoothness of the field.

The hard-core radius parameter is non-negative. Hence the exponential transformation on the field, $\mathbf{R}(s) := \exp(\mathbf{Y}(s))$, is employed, and this new random field is used in the Gibbs hard-core model. Let $[\mathbf{x}, \mathbf{R}]$ stand for the joint distribution of $\mathbf{x}$ and $\mathbf{R}$ as a short-hand notation. Then, the *varying range Gibbs hard-core model* with n points is defined as follows. The conditional model $[\mathbf{x}|\mathbf{R}]$ is defined by the conditional hard-core rule $\{\exists i : d_i < R(x_i)\}$ given $\mathbf{R}$. The associated conditional, heterogeneous total energy is then

$$U_{\mathbf{R}}(\mathbf{x}) = \sum_{i=1}^{n} \tilde{\phi}_{\mathbf{R}}(x_i, \mathbf{x}_{-i}) \quad \text{where} \quad \tilde{\phi}_{\mathbf{R}}(x_i, \mathbf{x}_{-i}) = \begin{cases} \infty, & \text{if } d_i < R(x_i) \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

This construction leads to a hierarchical two-level model, because $[\mathbf{x}, \mathbf{R}]$ can be written as $[\mathbf{R}][\mathbf{x}|\mathbf{R}]$. The observation model $[\mathbf{x}|\mathbf{R}]$ is an inhomogeneous point process model for any given smoothly varying $\mathbf{R}$.

2.3. Model synthesis

The model simulation is straightforward using the hierarchical presentation $[\mathbf{x}, \mathbf{R}] = [\mathbf{R}][\mathbf{x}|\mathbf{R}]$: One simulates the random field $\mathbf{Y}$, transforms $\mathbf{Y}(s) \rightarrow \mathbf{R}(s)$, and then simulates the Gibbs hard-core model with varying range parameter conditional on $\mathbf{R}$.

Random field generation

Simulation of GRFs faces the so called *big N* problem arising from the Cholesky decomposition of large, dense covariance matrices. It can be avoided using approximations based on the Central Limit Theorem, see e.g. Lantuéjoul (2002), which are implemented in the R package `RandomFields` (Schlather, 2011).

Here a computationally more efficient approximation, based on Markov random fields, is employed. Lindgren et al. (2011) prove that for some covariance functions in the Matérn family the Markov random field approximation is exact. This approximation together with sparse matrix algorithms discussed in Rue and Held (2005) and Lindgren et al. (2011) give a solution to the *big N* problem and provide quick computation needed in the Bayesian analysis later on. The approximation is done on a regular lattice in a rectangular window (details can be found in Appendix). The parameter field $\mathbf{R}$ is obtained from the GRF after one-to-one transformation.

Conditional simulation of the point process

Given a realisation $\mathbf{R}$ we then simulate the varying range Gibbs hard-core model with a fixed number of points using a combination of the birth-and-death (BD) algorithm (Ripley, 1977) and the Metropolis–Hastings (MH) algorithm (Møller and Waagepetersen, 2004). BD is advantageous when n, the number of points requested, is close to the maximal number of points that the random field $\mathbf{R}$ allows. "Birth" is an event, where the current (initially empty) pattern joined with a new proposal point agrees with the conditional hard-core rule. "Uniform death", i.e. removal of a randomly chosen point from the current pattern, is executed when a predefined number of consecutive failures to give birth has occurred. If the pattern reaches n points within some give-up iteration limit, it is considered as a starting configuration, not far from being an equilibrium state. It is then further relaxed using the MH algorithm, moving points to a randomly chosen new location and obeying the hard-core rule.

Simulation experiment

Fig. 2 depicts four realisations of the model with two different smoothness levels of the field. The left point pattern (middle column) is simulated so that the point count was approximately as high as the field permitted (give-up after $2 \cdot 10^6$

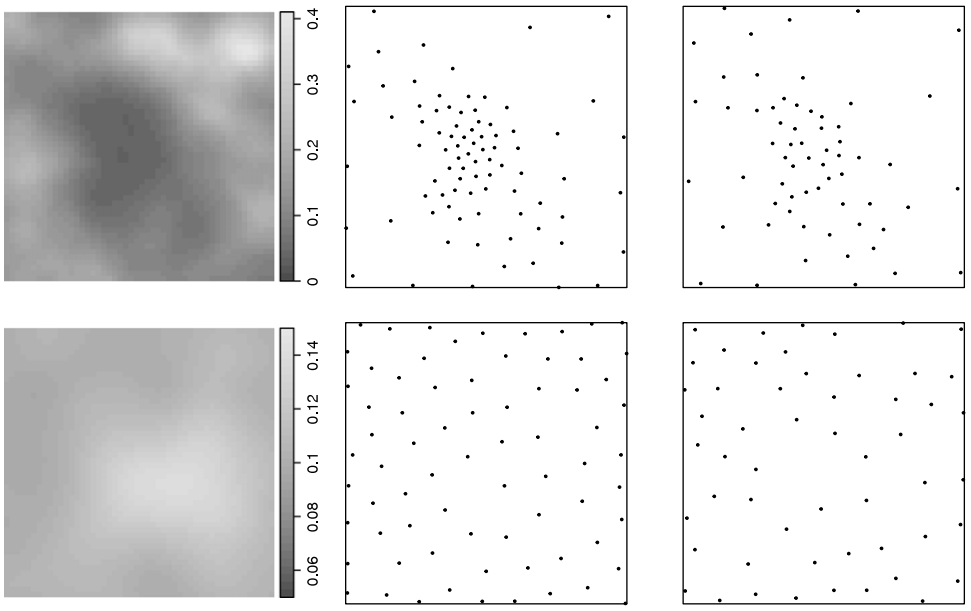

Fig. 2. Simulation of the varying range Gibbs hard-core model as described in Section 2.3. Top row: Latent field (top left) has a region of high values in the top right corner of the area, and the two simulations of the model (top middle, top right) exhibit low intensity there. Bottom row: The field is much smoother but still the intensity is not constant in the two simulation with this field. Patterns on the right have 75% of the maximal packing abundance approximately achieved in the left ones.

steps). The right pattern was simulated with 75% of the maximal count. The patterns reflect well the intended local scale repulsion and, much like the inhomogeneous Poisson model does, the spatial variation in intensity.

We explored also the possibility to apply perfect simulation of the initial pattern. Perfect simulation algorithms for point processes require a tractable lower limit to the energy function of Eq. (2) (Kendall and Møller, 2000; Cai, 2005; van Lieshout and Stoica, 2006). The resulting "dominating" Poisson process is then thinned within a coupled MCMC run. But in the hard-core models we would need the dominating starting sample from the set of (theoretically) maximally packed patterns, and this task is already in the case of constant R very demanding (Mase et al., 2001).

3. Bayesian analysis

In this section the likelihood (1) is refined and a Bayesian MCMC method is suggested to infer from the field $\mathbf{R}$ when the point pattern $\mathbf{x}$ is observed.

3.1. Intractable likelihood

Consider the inhomogeneous varying range hard-core Gibbs model with the total energy function (4) for a fixed $\mathbf{R}$. The normalising constant of the likelihood (1) is intractable, even for the homogeneous hard-core process. A popular approximation for homogeneous Gibbs models is the pseudolikelihood (PL) density (Besag, 1978; Jensen and Møller, 1991). However, the PL density requires the computation of the Papangelou intensity,

$$\lambda_{\mathbf{R}}(\mathbf{x}, s) \propto 1(\mathbf{x} \cup s \text{ is valid under } \mathbf{R}),$$

also in points s outside of $\mathbf{x}$ and this should be done within each MCMC step. In addition, the PL approximation is not necessarily satisfactory for (4) due to possibly large hard-core radii. Hence the PL approximation does not suffice and a new approach is required.

3.2. Modelling the nearest neighbour distances given $\mathbf{R}$ under the high packing point density assumption

Consider the Bayesian model with only the hard-core rule in the likelihood. Then the only local information on the hard-core radii would be that $R_i := R(x_i) \in [0, d_i]$ for all $i = 1, \ldots, n$. For large (unobserved) values of R_i, that is, for sparse parts of the observed point pattern, the posterior values of R_i tend to be towards the mean of the $\mathbf{R}$-field, allowing close pairs of points and invalidating the paradigm of high packing we have in mind. As a conclusion, the hard-core Gibbs model is too simple.

The hypothesis of high packing density imposes a strong connection between R_i and d_i. Suppose for a while that the hard-core model is homogeneous. Then the maximum likelihood estimator is given by the minimum of the observed nearest

neighbour distances $\{d_i\}$. This estimator is sometimes corrected by a factor $(n-1)/n$ since one is operating below the maximum packing density configuration, for which $d_i \approx R_i$. If the packing density is below maximal, one would expect a positive deviation or "slack" between R_i and d_i. This connection is analytically intractable.

Return now to the inhomogeneous hard-core model. Using simulation experiments we determined an empirical link between R_i and d_i. The first discovery is that it is better to work with the relative slack than with the absolute one, i.e. with the relative error instead of the additive one. The additive error can be written as a statistical model of the form

$$d_i - R_i = e_i, \quad e_i \sim p_{\theta_i},\ e \in [0, d_i].$$

The distribution p_{θ_i} is subject to changes from one point to another according to the parameters θ_i. Based on an extensive simulation trial (see Appendix) it was found that the distribution of *relative slack* $e_i := (d_i - R_i)/d_i$ is approximately spatially invariant. The distribution of relative slack, p_{θ_e}, is analytically intractable but empirically the single parameter power law

$$p_{\theta_e}(e) \propto (1-e)^{\theta_e - 1}, \quad e \in [0, 1],\ \theta_e > 1,$$

which equals Beta$(1, \theta_e)$, seems to be a reasonable approximation. Hence a single parameter θ_e is enough to model the slack over all points. Consequently, this approach is good when the proportional overhead in the repulsion area (e.g. territory) size d_i does not depend on the location nor on the absolute value R_i driving the process.

The density of $\mathbf{x}$ given $\mathbf{R}$ is now assumed to be the product of two components: the interaction potential and the slack. The first component controls the hard-core property and the second component controls the empty space behaviour. Furthermore, the slacks are assumed to be independent of each other. With these assumptions, the conditional likelihood conditional on $\mathbf{R}$ obtains the form

$$\begin{aligned} f(\mathbf{x}; \mathbf{R}, \theta_e) &\propto \exp\left\{-\sum_{i=1}^{n} \tilde{\phi}_{\mathbf{R}}(x_i, \mathbf{x}_{-i})\right\} \prod_{i=1}^{n} p_{\theta_e}(e_i) \\ &\propto \exp\left\{-\sum_{i=1}^{n} \tilde{\phi}_{\mathbf{R}}(d_i)\right\} \prod_{i=1}^{n} (1-e_i)^{\theta_e - 1} \\ &= \exp\left\{-\sum_{i=1}^{n} \hat{\phi}_{\mathbf{R}}(d_i)\right\} \end{aligned} \tag{5}$$

with

$$\hat{\phi}_{\mathbf{R}}(d_i) = \tilde{\phi}_{\mathbf{R}}(d_i) + (\theta_e - 1) \log\left(\frac{R_i}{d_i}\right),$$

which is a new member among the nearest neighbour Gibbs processes, see Bertin and Drouilhet (1999) and van Lieshout (2000). This model is capable of causing a "drift" towards higher packing in model fitting. The scaling factor is again intractable, and, because the distribution of the slack variable is not known, this should be determined empirically.

For the varying range hard-core Gibbs process the connection between the radii and relative slack can be studied by means of simulation. The slack distribution is highly dependent on two major assumptions in the model. The first is the unobserved packing density of the pattern. The difference between a "shoulder-to-shoulder" packing and "plenty-of-room" packing can be of factor 5 in the slack parameter. This feature provides a nice and important interpretation to the slack parameter θ_e. If we assume a dense packing, e.g. nearly full use of the environment's capacity to host individuals, we should assign θ_e a high value in order to have each R_i close to the observed d_i. Conversely, if we assume a relaxed pattern with plenty of extra available space for the potential individuals to utilise, we should set θ_e lower and allow more relative slack between the observed d_i and the model parameter R_i.

The second major factor related to the parameter θ_e is the smoothness of the field. Our empirical results show that the smoother the field, the more dependent the parameter is on the packing density of the pattern. This is illustrated in terms of a linear model between the packing density and the parameter θ_e, depicted in Fig. 3. With rough fields the packing density has a relatively low slope, but the slope increases when the smoothness increases. The linear fit is not perfect but works well as a guideline when deciding on the parameter, and thus on the packing density, in the initial modelling stages.

The maximum number of points a field can support also has an effect on the slack distribution. The local smoothness of the field is related to the intensity of the point pattern, so a rough field with high maximum number of points (lower mean repulsion ranges) is similar in local smoothness to a smooth field with low maximum number of points (higher mean).

3.3. Model fitting and diagnostics

As pointed out, smoothness of the random field $\mathbf{R}$ and the packing density are problematic because point pattern data contain very little information on these parameters. Furthermore, model diagnostic tools that rely on the posterior likelihood, such as DIC or Bayes factors, cannot be computed due to the intractable normalising constant. To overcome these problems the idea of *posterior predictive assessment* of model fitness is applied (Gelman et al., 1996). We fix the smoothness

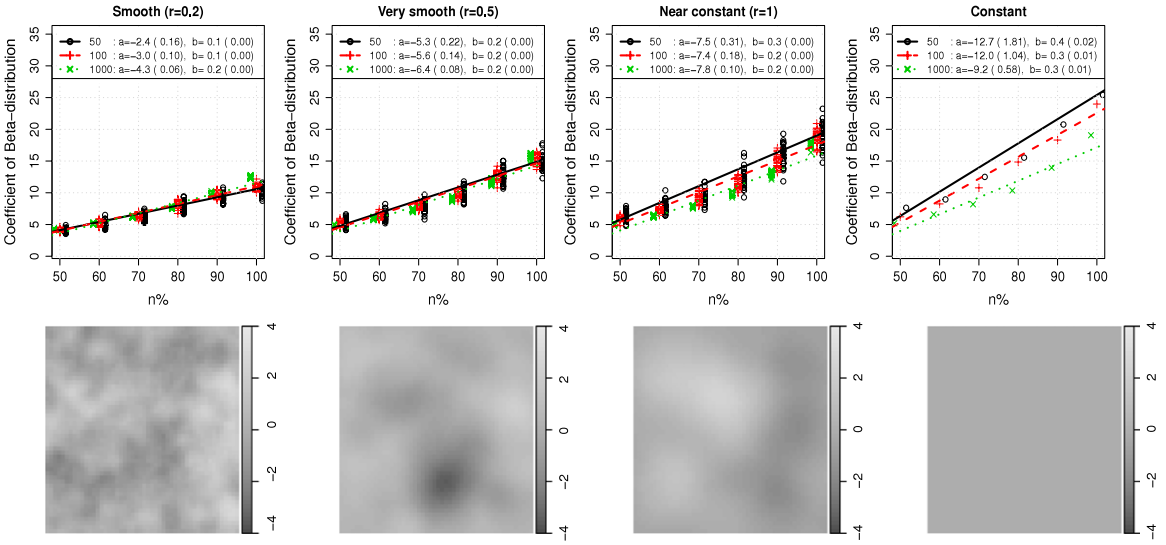

Fig. 3. Illustration of how the slack parameter depends on the assumed packing density and the smoothness of the latent field. Top: Linear model fits for three random fields and a constant field. Each plot depicts results separately for empirical maximal packing ($n\% = 100$) of ca. 50, 100 and 1000 points. A single symbol is the mean over 20 point pattern simulations, repeated over 25 field realisations. The regression line is based on the pooled 3000 patterns. Bottom: A standardised example of a GRF realisation of each smoothness class used in the simulation trial.

and packing density parameters, and run the MCMC fitting with the Gaussian prior and the density (5). The MCMC algorithm details are given in Appendix D. During the MCMC run, a posterior *discrepancy index* is computed. By repeating this for a reasonable combination of smoothness and packing density parameters a map of discrepancy indices are produced. This map can then be used to choose parameter values that give the best fitting model.

Predictive integrated mean square error (IMSE) of a summary field was used as the discrepancy index, with an adaptive kernel intensity estimator as the summary field. We scale the estimator with $|D|/n$ to facilitate comparison across datasets with different point counts (see details in Appendix B). After a burn-in period in the MCMC fitting algorithm we sampled from the posterior predictive distribution of $\mathbf{R}$ and then simulated a new point pattern $\mathbf{x}^\star$. Repeating the simulation K times during the MCMC run, K patterns are produced from the predictive distribution $[\mathbf{x}^\star|\mathbf{x}]$. For these simulations, a summary field MSE at grid cell location c_i is defined as

$$\mathrm{MSE}(c_i) := \mathrm{MSE}_i := \frac{1}{K}\sum_{j=1}^{K}(s_i^d - s_i^j)^2$$

where s_i^d is the summary field value for the data point pattern and s_i^j for the jth simulated point pattern. The IMSE is then an approximation to the integral of MSE over the area,

$$\mathrm{IMSE} := \int_D \mathrm{MSE}(x)dx \approx \sum_{i=1}^{N}\mathrm{MSE}_i\delta_i$$

with grid cell area δ_i and count N. A smaller value of IMSE indicates better fit.

The posterior samples of $\mathbf{R}$ from the IMSE-wise best MCMC run can then be used for posterior analysis, such as producing posterior mean and variance maps for $\mathbf{R}$, or maps for high packing areas defined as areas where R_i's are below some threshold with high posterior probability.

3.4. Simulation trial

We carried out a simulation experiment to examine how well the model is identified using the methods described in the previous sections.

We simulated a realisation of the GRF on a 50×50 grid in a $[0, 1]^2$ window with Matérn covariance and differentiability $\nu = 2$, range $r = 0.5$ and precision $\tau = 3$. The mean was set to $\mu = \log(0.1)$, and a pattern was simulated until no further points could be added, yielding a high packing density of 84 points. We call this the maximal packing density. Then on the same field, a second pattern was simulated with 75% of the maximal packing density, or 63 points. The field and the two patterns are shown on the top row of Fig. 2.

Fig. 4 depicts the results for fixed $\nu = 2$ and various beta and range parameters when the summary field is a scaled adaptive kernel intensity estimator. The range values that give best fit are between 0.3 and 0.5 which is close to the original 0.5. The slack parameter value is identified for the pattern near maximal (100%) packing (left pair in Fig. 4) to values around

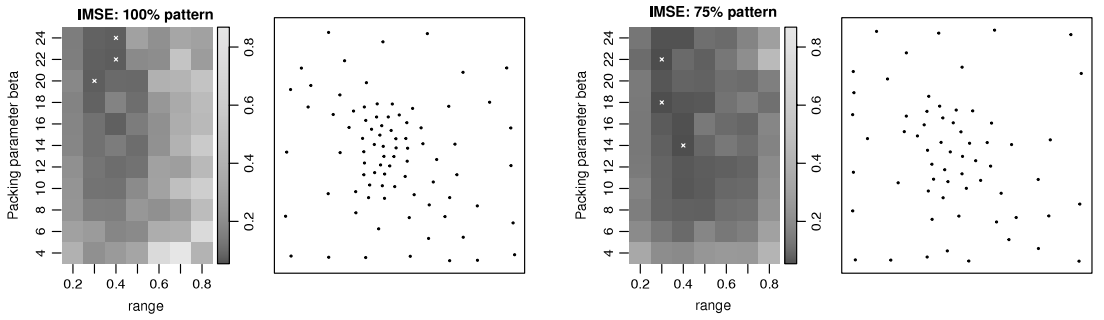

Fig. 4. Model fit as assessed by posterior predictive IMSE, and a single pattern realisation from the predictive distribution with the best fitting smoothness and packing density parameters. Darker colour represents smaller IMSE. Best three fits are marked with crosses in each IMSE map. Left pair is for the 100% packed and right pair is for the 75% packed pattern used in the simulation trial.

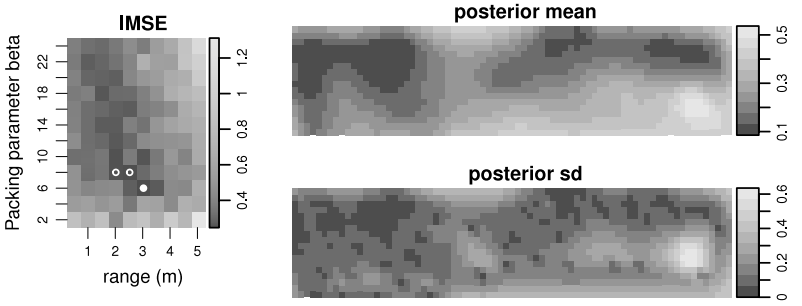

Fig. 5. Varying range hard-core model fitted for the Sand Martin data. Left: IMSE map for range and packing density parameter identification. Best three are marked with a circular symbol on the IMSE map, illustration example on the right uses the pair of parameter marked with the filled symbol. Right: Posterior mean and standard deviation of **R** field computed with the chosen parameter combination.

22 and for the sparser pattern (right pair) around 18. By the empirical regression results of Section 3.2 the posterior predictive assessment identifies correctly the 100% packing pattern but not the 75% packing.

Fig. 4 also shows realisations from the predictive distributions of the point processes with the best parameter combination in IMSE sense. The 100% prediction agrees well with the original pattern, and the 75% also depicts the 'valley' in the middle portion of the area.

4. Sand bank heterogeneity as observed through Sand Martin's nest locations

We return to the example in Section 1, Fig. 1, describing pattern of nests of Sand Martins in a vertical sand bank where the only information is the nest locations.

Sand Martins (*Riparia riparia*) are migratory passerine birds in the swallow family. They usually nest in colonies between 10 and 4000 pairs. The nests are holes in river banks, sandpits, claypits and other types of man-made excavations, and have been dug by the birds. An average hole measures 6 cm in width and 9 cm in height at the entrance, and is dug 20–200 cm deep into the bank at an average slope of 3°. Nest locations are strongly affected by the soil particle composition of the bank, smaller grain size being more favourable for burrowing into (Heneberg, 2003). The nest pattern tends to be highly packed in favourable parts of the sand bank due to shortage of suitable sand composition areas, competition on sites and the sub-colony structure in nest formation, see e.g. Sieber (1980) and Heneberg (2003).

The nest location data is derived from a photograph taken by H. Högmander in Konnevesi, Central Finland, in August 1988. After digitisation the photo was cropped to represent a 5.2 m $\times$ 1.3 m rectangular area with 67 nests (see the top plot of Fig. 1). A measure stick was included in the photo for scale. There is a vertical trend in the intensity, and the threat of nest cave-ins justifies the hard-core assumption.

Results are depicted in Fig. 5 when we fix $\nu = 2$. The range parameter identifies around 2–3 m, which is 40%–60% of the window width. The slack parameter identifies around 6–10. For illustration we assumed the pattern to be 75% from maximum packing, and chose parameters marked with a dot in Fig. 5. These slack-range values were also in the top 3 parameter values IMSE-wise. The posterior mean has low repulsion values at the top of the bank which indicates smaller grain size and sand composition good for nest burrowing.

A predictive point pattern simulation and the mean predictive intensity field from the chosen model is shown in the bottom plot of Fig. 1. The shape of the field is close to that of the original data, and the simulation depicts the hard-core effect as intended.

With a sample of the bank's grain composition we could interpret the field values using the results in Heneberg (2003).

5. Discussion

The problem of inhomogeneity in the form of varying hard-core radius in a spatial point pattern in considered. The seemingly simple question turned out to be very complicated because the information on the underlying Gaussian process is not transformed to the posterior due to the lack of local information outside data. The posterior should be penalised towards higher packing density.

As a solution, the local nearest neighbour distances for each observed point in the observed point pattern is included into the new model. This model is a variant of the nearest-neighbour Gibbs process.

A Markov approximation (Rue and Held, 2005; Lindgren et al., 2011) was used to avoid the *big N* problem of GRFs. Given a realisation of the latent field, the simulation of the point pattern used the Metropolis–Hastings algorithm with an initial birth-and-death algorithm step to acquire the starting hard-core pattern. Statistical inference used the nearest neighbour approximated likelihood and the Markov approximation of GRF inside the update step in MCMC posterior simulation. The algorithm produced predictions of the hard-core radius field outside data points and of the conditional point process itself for model diagnostics.

Several possibilities of improving and generalising the model are possible. First, the underlying latent GRF could be anisotropic which would also improve the analysis of Sand Martin's nest data because often the sand structure in sand banks has horizontal layers.

Secondly, the most efficient way of handling inhomogeneity is to utilise available covariates in the model. In the Sand Martin nest data such biological covariates would be height of a nest location from the top and bottom of the bank, to describe predator aversion, and small scale territorial repulsion (Sieber, 1980). Also measurements of the sand grain sizes would predict the nest locations. If covariates are available, these could be applied for the first order term in the Gibbs energy function (Stoyan and Stoyan, 1994; Baddeley and Turner, 2000) in the spirit of generalised linear modelling, or the covariates could be used in the mean of the GRF for the pairwise interactions. It is an open question how well covariates would penalise the latent hard-core random field towards higher packing density.

Acknowledgments

The work is part of the Ph.D. studies of the first author who is funded by COMAS and FDPSS doctoral programs. We thank H. Högmander for providing the Sand Martin data at our disposal. Our thanks are also due to the Associate Editor and two anonymous reviewers for their valuable comments.

Appendix A. Markov approximation of the GRF

Assume D is rectangular. We do our approximative computations on a finite grid G with cell centers $\{c_k : c_k = (x_i, y_j), i = 1, \dots, n_x, j = 1, \dots, n_y\}$ and with the partition $\bigcup_{i,j}[x_i - \frac{\delta_x}{2}, x_i + \frac{\delta_x}{2}] \times [y_j - \frac{\delta_y}{2}, y_j + \frac{\delta_y}{2}] = D$ where δ_x and δ_y are the cell diameters in x- and y-coordinate directions, respectively. We assume $\delta_x = \delta_y$ i.e. square cells. Write $N = n_x n_y$, and $\delta_k = \delta_x \delta_y$.

Details of the Markov field approximation of GRF are given in Rue and Held (2005). The explicit link between the Matérn covariance family and their Markov approximation are given by Lindgren et al. (2011). In short, the GRF is approximated on a grid G using a local cell neighbourhood that depends on the differentiability parameter ν. Markovianity makes most of the precision matrix $Q = \Sigma^{-1}$ values zero. Then the sparseness of Q is made use of in fast computation of the Cholesky factorisation utilised in the Gaussian density.

In our examples the point patterns were relatively small, so a (dense matrix) multivariate normal model was applied when only the data locations were involved. Most parts of the MCMC algorithm were just that. A conditional prediction outside data locations was produced by setting the data values to their closest grid points, conditioning on these grid points, and using a Markov approximation in the conditional simulation. We do the Markov approximation also when a realisation of the Gaussian field is required, as in the simulation of the process.

Appendix B. Adaptive intensity estimator

The fixed bandwidth kernel estimator for a heterogeneous intensity field $\lambda(x)$, computed on the location of a finite grid with a Gaussian kernel $\phi(r) = \frac{\exp(-0.5r^2)}{2\pi}$ and some bandwidth $h > 0$, is

$$\lambda(x) \approx \tilde{\lambda}(x) = \frac{1}{n}\sum_{i=1}^{n} \phi(\|x_i - x\|/h)h^{-2}$$

with point pattern data $\mathbf{x} = \{x_i : i = 1, \dots, n\}$. Following Silverman (1986) and Diggle et al. (2005) we extend this to an adaptive intensity estimator by plugging in a location dependent bandwidth, computed using a pilot kernel estimate with fixed bandwidth estimator, i.e.

$$\hat{\lambda}(x) = \frac{1}{n}\sum_{i=1}^{n} \phi(\|x_i - x\|/h(x_i))h(x_i)^{-2}, \quad \text{where } h(x) = \frac{h_0}{\sqrt{\tilde{\lambda}(x)\tilde{\gamma}}}.$$

Table C.1
Parameters for different GRFs used in the link distribution trial.

Descriptive name in Fig. 3	Rough[a]	Smooth	Very smooth	Near constant
Precision τ	5	10	10	10
Range r	0.1	0.2	0.5	1
Differentiability ν	2	2	3	3

[a] Not shown.

The scaling factor $\tilde{\gamma}$ is the geometric mean of the estimates $\tilde{\lambda}$ over the grid. The global bandwidth h_0 was set using the overdispersion parameter as described in Davies and Hazelton (2010). Bias near the border was countered following Diggle et al. (2005).

To facilitate IMSE comparisons across datasets with different point counts n and different domains D, we scale the adaptive intensity estimator with $|D|/n$.

Appendix C. Details of the slack distribution simulation trial

A simulation trial was done to study the link between true field values R_i and the nearest neighbour distances d_i for a point pattern $\mathbf{x} = \{x_i\}$ simulated from the varying range Gibbs model given $\mathbf{R}$. We want to build a statistical model between the observed point pattern $\mathbf{x}$ and the latent, unobserved field $\mathbf{R}$.

In total we fixed six different parameter sets $\theta := (\tau, r, \nu)$ to produce the Matérn field realisation with different smoothnesses, ranging from independent noise to a constant field. We do not show the noise field or the roughest dependent field in this paper as the more smoother ones are more realistic in our application. The precision for noise was 4, and for the other stochastic fields as in Table C.1.

For each parameter set $\theta^{(s)}$, $s = 1, \ldots, 6$, we simulated 25 GRF realisations $\mathbf{Y}_i^{(s)}$, $i = 1, \ldots, 25$, using the Markov field approximation as discussed in Appendix A. Exceptions were the noise field with 5 simulations and the constant field with no need for simulation. By default the window $D = [0, 1]^2$. The fields have mean 0 at this stage.

To study whether the total point count at maximum packing capacity affects the slack distribution we chose three maximum point count 'classes' for comparison: $N_j = 50, 100, 1000$. Since the maximum packing capacity varies between field realisations, an estimate of the empirical maximum point count in each individual field realisation is needed. We therefore simulated 10 point patterns on each field with the target number of points set to 10 000, and by trial and error we adjusted the field means $\mu_{s,i,j}$ so that each exp-transformed field $\mathbf{R}_{i,j}^{(s)}(c) = \exp(\mathbf{Y}_i^{(s)}(c) + \mu_{s,i,j})$, $c \in G$, could facilitate about N_j points. The maximum numbers per field realisation, $N_{i,j}^{(s)}$, were recorded. The maximum numbers were classified to 6 density classes $n_{i,j,t}^{(s)} = tN_{i,j}^{(s)}$ with the factor $t = 0.5, 0.6, \ldots, 0.9, 1.0$ corresponding to percentual packing densities ranging from very sparse (50%) to fully packed (100%).

Then, finally, for each level of smoothness s, individual field realisation i, empirical maximum abundance j and relative packing density t, 20 point patterns $x_{i,j,t,k}^{(s)}$, $k = 1, \ldots, 20$ were simulated. From each pattern we collected and stored the nearest neighbour distance d_i and the field value R_i of each point. To counter border effects we discarded all values of points with distance to border less than the nearest neighbour distance. The slack term $e_i = (d_i - R_i)/d_i$ for each point was computed and analysed in order to determine a suitable statistical model.

Appendix D. Bayesian MCMC algorithm description

Grid resolution was set so that each grid cell held at most one data point. In simulation examples a 50×50 grid was sufficient. Write $O \subseteq \{1, \ldots, N\}$ for cells with data and $S = \{1, \ldots, N\} \setminus O$ for cells where we want to predict the field values using data. Write $\mathbf{R}_o = \{R_k : k \in O\}$ and $\mathbf{R}_s = \{R_k : k \in S\}$.

The algorithm was designed to estimate the mean parameter μ, the precision parameter τ and the field values $\mathbf{R}_o$, to provide predictions of the field values $\mathbf{R}_s$ using conditional simulation, and to sample from the predictive distribution of $[\mathbf{x}^\star|\mathbf{x}]$ for model diagnostics.

The following parameters were considered fixed by the user: range r and differentiability parameter ν of the field; the packing density parameter β. Other user parameters of the algorithm are: number of $\mathbf{R}_o$'s to update per iteration; size of grid expansion for border correction; random seed; interval for prediction $\mathbf{R}_s$; interval for prediction $\mathbf{x}^\star$ (includes prediction of $\mathbf{R}_s$); verbosity.

Note that data have been used only for modelling $\mathbf{R}_o$. $\mathbf{R}_s$ is treated as a conditional simulation problem given $\mathbf{R}_o$, and therefore computed only when required.

The original model is

$$x_i|\mathbf{R} : \ \mathbf{x} \sim f(\mathbf{x}; \mathbf{R}) \quad \text{(heterogenous Gibbs hard-core)}$$
$$R(x_i)|\tau, \mu : \ \log(\mathbf{R}) \sim \mathrm{GRF}(\mu, \tau, C_{r,\nu})$$

with r, ν fixed. We use a Gaussian prior for the mean and a Gamma prior for the precision.

Due to the intractability and PL accuracy problems discussed in the text, the heterogeneous Gibbs hard-core likelihood is replaced with a statistical sampling model for the nearest neighbour distances, i.e. $x_i|\mathbf{R}$ is replaced by

$$d_i|R_i : \frac{d_i - R_i}{d_i} \sim \text{Beta}(1, b).$$

The hard-core rule can be ignored as due to the assumption $R_i \in [0, d_i]$ it is always one. Then the posterior model can be written as

$$p(\mathbf{R}, \tau, \mu|\mathbf{x}, r, \nu, b) \propto \left(\prod_{i=1}^{n} p_b(d_i|R_i)\right) p_{r,\nu}(\mathbf{R}_o|\tau, \mu)p(\tau)p(\mu).$$

Given the state at iteration t, $(\mu^{(t)}, \tau^{(t)}, \mathbf{R}_o^{(t)})$, the algorithm takes the following steps:

1. Update $\mu^{(t)}$ using a Gibbs sampler step.
2. Update $\mathbf{R}_o^{(t)}$ using a MH-step.
3. Update $\tau^{(t)}$ using a MH-step.
4. If requested, simulate R_s and/or prediction $\mathbf{x}^\star$.

Gibbs sampler step: With μ_0, τ_0 as hyper parameters for μ, posterior of μ is Gaussian with mean and precision

$$\mu_{\mu'} = \frac{B}{A}, \qquad \tau_{\mu'} = A, \quad \text{where} \quad A = \tau_0 + \sum_{i,j\in O} Q_{ij},\ B = \tau_0\mu_0 + \frac{1}{2}\sum_{i,j\in O} Q_{ij}(\log R_i + \log R_j).$$

Update $\mathbf{R}_o^{(t)}$: Proposition are made by first sampling the beta-distribution $e_i' \sim \text{Beta}(1, b)$ and then transforming to R_i'. A subset $I \subseteq O$ of indices is updated to maintain 10%–20% acceptance ratio. New values do not violate hard-core rule since we constrain each new value to $R_i' \in [0, d_i]$. Acceptance ratio is

$$\text{MH}(\mathbf{R}_o'|\mathbf{R}_o) = \exp\left(-\frac{\tau}{2}\left[(\log\mathbf{R}_o' - \underline{\mu})^T Q_o(\log\mathbf{R}_o' - \underline{\mu}) - (\log\mathbf{R}_o - \underline{\mu})^T Q_o(\log\mathbf{R}_o - \underline{\mu})\right]\right)\prod_{i=1}^{n}\frac{R_i}{R_i'}$$

with $\underline{\mu} = (\mu, \ldots, \mu)^T$ the same length as R_o.

Update $\tau^{(t)}$: Proposition is $\tau' = f\tau$, where $f > 0$ has a Lambert distribution. The transition probability is then symmetric (Rue and Held, 2005, p. 146). The MH ratio is given by

$$\text{MH}(\mathbf{R}_o'|\mathbf{R}_o) = \left(\frac{\tau'}{\tau}\right)^{N/2} \exp\left(-\frac{\tau' - \tau}{2}(\log\mathbf{R}_o - \underline{\mu})^T Q_o(\log\mathbf{R}_o - \underline{\mu})\right)\frac{p(\tau')}{p(\tau)}$$

where $p(\tau)$ is the prior.

Our experience is that the algorithm requires an informative prior for τ. It has an effect on the posterior variance on the latent field, but the posterior mean is more robust against the prior choice.

Appendix E. Details of the simulation trial

τ-prior selection was based on an empirical variogram estimate where the data value at x_i was d_i. Gamma(1.5, 0.5) was fixed for all fits. Note that it is common to use an informative prior as the field has very little information on the precision. Smoothness was fixed because its simultaneous estimation with precision is inconsistent (Zhang, 2004); see also the discussion in Section 3.3.

Each range-beta-fit pair had 10 000 iterations and predictive distribution was sampled every 100th step. After 5000 steps for burn-in and with 200 step thinning, the sample size of predictive simulations used in goodness-of-fit evaluations was $K = 25$.

For each fit an adaptive intensity field was computed using Gaussian kernel on a 50×50 grid ($N = 2500$). The IMSE was computed as described in Section 3.4.

Appendix F. Software

Simulation algorithm for GRF was implemented in R. Sparse matrix computations were done with the Matrix-package of R (Bates and Maechler, 2011).

Simulation algorithm for the point process given a realisation of GRF was implemented in C++ and used through R.

An R-package that implements the GRF approximation, the point pattern simulation, the MCMC algorithm and the computational algorithms discussed in Section 3.3 and Appendix B is available from the first author.

Appendix G. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.csda.2012.08.014.

References

Baddeley, A., Turner, R., 2000. Practical maximum pseudolikelihood for spatial point patterns. Aust. N. Z. J. Stat. 42, 283–322.
Bates, D., Maechler, M., 2011. Matrix: sparse and dense matrix classes and methods. R Package Version 0.999375-50.
Berthelsen, K.K., Møller, J., 2008. Non-parametric Bayesian inference for inhomogeneous Markov point processes. Aust. N. Z. J. Stat. 50, 257–272.
Bertin, E., Drouilhet, R., 1999. Existence of 'nearest-neighbour' spatial Gibbs models. Adv. Appl. Probab. 31, 859–909.
Besag, J., 1978. Some methods of statistical analysis for spatial data. Bull. Int. Stat. Inst. 47, 77–92.
Bognar, M., 2005. Bayesian inference for spatially inhomogeneous pairwise interacting point processes. Comput. Statist. Data Anal. 49, 1–18.
Cai, Y., 2005. A non-monotone CFTP perfect simulation method. Statist. Sinica 15, 927–943.
Cressie, N., 1993. Statistics for Spatial Data. Wiley.
Davies, T., Hazelton, M., 2010. Adaptive kernel estimation of spatial relative risk. Stat. Med. 29, 2423–2437.
Diggle, P., Rowlingson, B., Su, T., 2005. Point process methodology for on-line spatio-temporal disease surveillance. Environmetrics 16, 423–434.
Gelfand, A., Diggle, P., Fuentes, M., Guttorp, P., 2010. Handbook of Spatial Statistics. CRC Press.
Gelman, A., Meng, X.L., Stern, H., 1996. Posterior predictive assessment of model fitness via realized discrepancies. Statist. Sinica 6, 733–759.
Guttorp, P., Thorarinsdottir, T.L., 2011. Bayesian inference for non-Markovian point processes. In: Porcu, E., Montero, J.M., Schlather, M. (Eds.), Space–Time Processes and Challenges Related to Environmental Problems. Springer.
Hahn, U., Vedel Jensen, E.B., van Lieshout, M.C., Nielsen, L.S., 2003. Inhomogeneous spatial point processes by location-dependent scaling. Adv. Appl. Probab. 35, 319–336.
Heneberg, P., 2003. Soil particle composition affects the physical characteristics of Sand Martin *Riparia riparia* holes. Ibis 145, 393–399.
Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns. Wiley.
Illian, J., Sørbye, S.H., Rue, H., A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation. In: INLA. Ann. Appl. Stat. (in press).
Jensen, J., Møller, J., 1991. Pseudolikelihood for exponential family models of spatial point processes. Ann. Appl. Probab. 1, 445–461.
Kendall, W.S., Møller, J., 2000. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. Adv. Appl. Probab. 32, 844–865.
Lantuéjoul, C., 2002. Geostatistical Simulation: Models and Algorithms. Springer-Verlag.
Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B 74.
Månsson, M., Rudemo, M., 2002. Random patterns of nonoverlapping convex grains. Adv. Appl. Probab. 34, 718–738.
Mase, S., Møller, J., Stoyan, D., Waagepetersen, R.P., Döge, G., 2001. Packing densities and simulated tempering for hard core Gibbs point processes. Ann. Inst. Statist. Math. 53, 661–680.
Møller, J., Pettitt, A., Reeves, R., Berthelsen, K., 2006. An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. Biometrika 93, 451–458.
Møller, J., Waagepetersen, R.P., 2004. Statistical Inference and Simulation for Spatial Point Processes. Chapman & Hall, CRC.
Nielsen, L.S., Vedel Jensen, E.B., 2004. Statistical inference for transformation inhomogeneous point processes. Scand. J. Stat. 31, 131–142.
Rasmussen, C.E., Williams, C., 2006. Gaussian Processes. The MIT Press.
Ripley, B.D., 1977. Modelling spatial patterns. J. R. Stat. Soc. Ser. B 39, 172–212.
Rue, H., Held, L., 2005. Gaussian Markov Random Fields. Chapman & Hall, CRC.
Sampson, P.D., Guttorp, P., 1992. Nonparametric estimation of nonstationary spatial covariance structure. J. Amer. Statist. Assoc. 87, 108–119.
Schlather, M., 2011. Random fields: simulation and analysis of random fields.
Sieber, O., 1980. Kausale und funktionale Aspekte der Verteilung von Uferschwalbenbruten (*Riparia riparia* L.). Z. Tierpsychol. 52, 19–56.
Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. In: Monographs on Statistics and Applied Probability, Chapman & Hall.
Stoyan, D., Kendall, W.S., Mecke, J., 1995. Stochastic Geometry and its Applications, second ed. Wiley.
Stoyan, D., Stoyan, H., 1994. Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics. John Wiley and Sons.
van Lieshout, M.N.M., 2000. Markov Point Processes and their Applications. Imperial College Press.
van Lieshout, M.N.M., Stoica, R., 2006. Perfect simulation for marked point processes. Comput. Statist. Data Anal. 51, 679–698.
Vedel Jensen, E.B., Nielsen, L.S., 2000. Inhomogeneous Markov point processes by transformation. Bernoulli 6, 761–782.
Zhang, H., 2004. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. J. Amer. Statist. Assoc. 99, 250–261.