RESEARCH REPORT
www.csgb.dk

2010

Jesper Møller and Jakob Gulddahl Rasmussen

A Sequential Point Process Model and Bayesian Inference for Spatial Point Patterns with Linear Structures

A Sequential Point Process Model and Bayesian Inference for Spatial Point Patterns with Linear Structures

Jesper Møller & Jakob Gulddahl Rasmussen
Department of Mathematical Sciences
Aalborg University, Fredrik Bajersvej 7E, DK-9220 Aalborg
email: jm@math.aau.dk, jgr@math.aau.dk *

July 22, 2010

Abstract

We introduce a flexible spatial point process model for spatial point patterns exhibiting linear structures, without incorporating a latent line process. The model is given by an underlying sequential point process model, i.e. each new point is generated given the previous points. Under this model the points can be of one of three types: a 'background point', an 'independent cluster point', or a 'dependent cluster point'. The background and independent cluster points are thought to exhibit 'complete spatial randomness', while the conditional distribution of a dependent cluster point given the previous points is such that the dependent cluster point is likely to occur closely to a previous cluster point. We demonstrate the flexibility of the model for producing point patterns with linear structures, and propose to use the model as the likelihood in a Bayesian setting when analyzing a spatial point pattern exhibiting linear structures but where the exact mechanism responsible for the formations of lines is unknown. We illustrate this methodology by analyzing two spatial point pattern data sets (locations of bronze age graves in Denmark and locations of mountain tops in Spain) without knowing which points are background points, independent cluster points, and dependent cluster points.

Keywords: clustering; Dirichlet tessellation; simulation-based Bayesian inference; spatial point process.

1 Introduction

Many observed spatial point patterns contain points placed roughly on line segments, as exemplified by the locations of barrows shown in Figures 1a and the locations of mountain tops shown in 1b (further details on these data sets are given in Section 2.3). Such linear structures may be associated to a line segment process. For example, in Berman (1986) (see also Berman and Diggle (1989) and Foxall and Baddeley (2002)) both a point process representing copper deposits and a line segment process representing geological faults are observed and investigated for a possible association. Another example is animal latrines near territory boundaries modelled by the edges of an unobserved (deformed) Dirichlet (or Voronoi) tessellation (Blackwell 2001, Blackwell and Møller 2003, Skare, Møller and Jensen 2007). However, in many other cases, including the two data sets in Figure 1, the exact mechanism responsible for the formations of lines is unknown. Thus the development of tractable and practically useful spatial point process models, capable of producing point patterns with linear structures—but without introducing an observed or latent line segment process—becomes important.

———

Figure 1 about here

———

In this paper we introduce a flexible point process model with linear structures without incorporating any line segment process into the model. It is a sequential point process model, i.e. the points has an ordering, and moreover the points can be of one of three types: a 'background point', an 'independent cluster point', or a 'dependent cluster point'. The background and independent cluster points are thought to exhibit 'complete spatial randomness', while the conditional distribution of a dependent cluster point given the previous points is such that the dependent cluster point is likely to occur closely to a previous cluster point. The model turns out to be easy to simulate and interpret. Section 2 provides further details on the model and shows simulated realizations.

The joint density for our sequential point process model turns out to be expressible in closed form and depends on an unknown parameter $\theta = (p, q, \sigma) \in [0, 1] \times [0, 1] \times (0, \infty)$. Briefly, $1 - q, (1 - p)q, pq$ specify the probabilities of the three types a point can be ('background', 'independent cluster', or 'dependent cluster' point), while σ controls the spread of a dependent cluster point around the previous cluster points. However, we assume that only the collection of the points is observable within a given bounded planar region W (the observation window) meaning that both the types and the ordering of the points are unknown. Section 3 proposes to treat this 'missing data problem' by using a simulation-based Bayesian approach with a prior on θ and incorporating the missing data into the posterior. The usefulness of this methodology is illustrated by analyzing the two data sets in Figure 1.

2 The sequential point process model

Let $W \subset \mathbb{R}^2$ be a given bounded region of area $|W| > 0$. We think of W as an observation window, such as the rectangular regions in Figure 1, and we suppose that a finite point pattern $\{x_1, \ldots, x_n\} \subset W$ is observed. For convenience, when introducing the model for cluster points in Section 2.1, we assume that W is convex, but it is possible to extend our setting to the case where W is not convex.

In the vast literature on spatial point processes, most models are specified directly as a Poisson process, a Cox process (i.e. a hierarchical model for a Poisson process conditional on a random intensity function), a Poisson cluster process, or by a density function which often is of the Markov or Gibbs type, see Møller and Waagepetersen (2004), Møller and Waagepetersen (2007), and the references therein. However, the recognition of an underlying space-time point process may not only be natural but also lead to a more tractable model. A particular simple and useful case is sequential point process models, i.e. sequential constructions with an ordering of the points. Such models include in particular Matérn hard core processes of types II and III and more generally random sequel adsorption models, see Møller, Huber and Wolpert (2010) and the references therein, but our particular sequential point process model introduced in the sequel seems new.

For simplicity and because of lack of information in the applications we have in mind, we ignore modelling the distribution of both n and the waiting times between point occurrences, i.e. we fix n to be the observed number of points, and we consider a setting with discrete times. We let $x_c = (x_1, \ldots, x_k)$ and $x_b = (x_{k+1}, \ldots, x_n)$ denote the ordered cluster points and ordered background points, respectively, where k is random. Note that these orderings concern not all the n points together, so we only have that x_1 and x_k are the first and last cluster points, and x_{k+1} and x_n the first and last background points. First the model for x_c conditional on k is specified in Section 2.1, and second the full model for (x_c, x_b) is specified in Section 2.2 such that we do not need to take the ordering of all the n points into consideration. Note that both the subpatterns $\{x_1, \ldots, x_k\}$ and $\{x_{k+1}, \ldots, x_n\}$ and their orderings are unknown, since only $\{x_1, \ldots, x_n\}$ is observed. This missing data problem is first treated in Section 3.

2.1 The model for cluster points

We make the following model assumptions for the cluster points, where the number of cluster points $k > 0$ is fixed, and $p \in [0, 1]$ and $\sigma > 0$ are unknown parameters whose meaning will soon be clear. The first cluster point x_1 is uniformly distributed on W, with density

$$f(x_1) = 1/|W|, \quad x_1 \in W.$$

For $i = 2, \ldots, k$, conditional on the cluster points $x_1, \ldots, x_{i-1}$, with probability p, the next cluster point x_i becomes a dependent cluster point, with a density $h(x_i | \{x_1, \ldots, x_{i-1}\}; \sigma)$ with respect to Lebesgue measure

on W and specified in (2) below; otherwise x_i becomes an independent cluster point, which is uniformly distributed on W (and hence independent of the previous cluster points). So x_i conditional on $x_1, \ldots, x_{i-1}$ follows the mixture density

$$f(x_i|x_1, \ldots, x_{i-1}; p, \sigma) = p h(x_i|\{x_1, \ldots, x_{i-1}\}; \sigma) + (1-p)/|W|, \quad x_i \in W. \tag{1}$$

For notational convenience, if $i = 1$, we set $f(x_i|x_1, \ldots, x_{i-1}; p, \sigma) = f(x_1)$.

As the notation indicates, $h(x_i|\{x_1, \ldots, x_{i-1}\}; \sigma)$ will not depend on the ordering of the previous cluster points $x_1, \ldots, x_{i-1}$. Specifically, for $i > 1$ and $x_1, \ldots, x_i \in W$,

$$h(x_i|\{x_1, \ldots, x_{i-1}\}; \sigma) = \frac{l_i^2 \exp(-r_i^2/\lambda)}{\lambda|W|(1 - \exp(-l_i^2/\lambda))} \quad \text{if } 0 < r_i < l_i, \tag{2}$$

and $h(x_i|\{x_1, \ldots, x_{i-1}\}; \sigma) = 0$ otherwise, where $\lambda = 2\sigma^2$, and r_i and l_i are defined as follows. Consider the Dirichlet (or Voronoi) tessellation of $\mathbb{R}^2$ with nuclei $x_1, \ldots, x_{i-1}$, that is the tessellation with cells

$$C_j = \{\xi \in \mathbb{R}^2 : \|\xi - x_j\| \le \|\xi - x_{j'}\|, j' = 1, \ldots, i-1\}, \quad j = 1, \ldots, i-1,$$

where $\|\cdot\|$ denotes Eucledian distance (see e.g. Okabe, Boots and Sugihara (1992) or Møller (1994)). Note that $C_j \cap W$ is convex—this is where the assumption that W is convex becomes convenient—and x_i belongs almost surely to a unique Dirichlet cell, say C_j. Then $r_i = \|x_i - x_j\|$ and l_i is the length of the line segment through x_i and with endpoints at x_j and the boundary of $C_j \cap W$. See the example in Figure 2a, where $i = 4$ and $j = 2$. Making a change of the variable $x_i - x_j$ to the polar coordinates (θ_i, r_i), we obtain

$$\int_{C_j} \frac{l_i^2 \exp(-r_i^2/\lambda)}{\lambda|W|(1 - \exp(-l_i^2/\lambda))} \, \mathrm{d}x_i = \int_0^{2\pi} \left(\int_0^{l_i} \frac{r_i l_i^2 \exp(-r_i^2/\lambda)}{\lambda|W|(1 - \exp(-l_i^2/\lambda))} \, \mathrm{d}r_i \right) \mathrm{d}\theta_i = \int_0^{2\pi} \frac{l_i^2}{2|W|} \, \mathrm{d}\theta_i = \frac{|C_j|}{|W|}.$$

Consequently, since $|W| = \sum_{j=1}^{i-1} |C_j \cap W|$, (2) is indeed a density function with respect to Lebesgue measure restricted to W. It also follows that x_i appears in cell C_j with probability $p_j = |C_j \cap W|/|W|$, and if x_i is conditioned to be in C_j, then x_i follows the restriction of $N_2(x_j, \sigma^2 I)$ to $C_j \cap W$, where $N_2(x_j, \sigma^2 I)$ denotes the radially symmetric bivariate normal distribution with mean x_j and standard deviation σ. This justifies the terminology 'dependent cluster point', but it is perhaps surprising that realizations of the model for cluster points exhibit linear structures as demonstrated later on.

———

Figure 2 about here

———

As explained in the following two paragraphs, an important advantages of the model is that neither the calculation of the distribution $p_1, \ldots, p_{i-1}$ or the construction of the entire Dirichlet tessellation is needed when evaluating the density (2) or simulating from this distribution.

To evaluate the density (2) we use the following simple steps, where we exploit the construction of Dirichlet cells and the fact that $C_j \cap W$ is convex for all $j < i$.

(a) Find the closest point x_j to x_i with $j < i$, the half-line L_j with endpoint x_j and passing through x_i, and the intersection point v_j between L_j and the boundary of W. Calculate $l_j = \|v_j - x_j\|$.

(b) For each $j' \in \{1, \ldots, i-1\} \setminus \{j\}$, find the line $L_{j'}$ passing through $(x_j + x_{j'})/2$ and perpendicular to the line through x_j and $x_{j'}$. If $v_{j'}$ is the intersection point between L_j and $L_{j'}$ within W, then calculate $l_{j'} = \|v_{j'} - x_j\|$. If $L_j \cap L_{j'} \cap W = \emptyset$, then set $l_{j'} = \infty$.

(c) Return $r_i = \|x_i - x_j\|$ and $l_i = \min\{l_1, \ldots, l_{i-1}\}$.

Figure 2b shows an example, where $i = 5$, step (a) returns $j = 4$ and $l_4 = \|v_4 - x_4\|$, step (b) returns $l_1 = \|v_1 - x_4\|$, $l_2 = \infty$, and $l_3 = \|v_3 - x_4\|$, and step (c) returns $r_5 = \|x_5 - x_4\|$ and $l_5 = \|v_3 - x_4\|$. Since $C_4 \cap W$ is the region around x_4 bounded by the lines L_1, L_2, L_3 and the boundary of W, (c) returns the correct result.

Moreover, we can easily make a simulation under (2) by the following steps.

3

(A) Generate a uniform point y_i in W, which is independent of $x_1, \ldots, x_{i-1}$.

(B) Find the (almost surely unique) closest point x_j to y_i ($1 \le j < i$), the half-line L_j through y_i and with end-point x_j, and the distance l_i from x_j to the intersection point between L_j and the boundary of $C_j \cap W$ (we calculate l_i in the same way as in (a)-(c) above).

(C) Generate t_i from an exponential distribution with parameter $1/\lambda$ and restricted to the interval $(0, l_i^2)$. Set $r_i = \sqrt{t_i}$.

(D) Return x_i as the point on L_j with distance r_i from x_j.

Figure 3a shows a simulation of x_c when W and $k = 81$ are as in Figure 1b (the mountain tops), and $p = 0.800$ and $\sigma = 286.4$ m are given by the estimated posterior mean obtained by the Bayesian method in Section 3 (with $q = 1$) using the mountain tops data set. Nearly all of the points in Figure 3a are placed in medium sized linear structures, whereas the point pattern in Figure 1b shows larger linear structures but also solitary points. As shown in Sections 2.3.2 and 3, the extension in Section 2.2 of the model to the case with the background points included provides a much better description of the mountain tops data set.

———

Figure 3 about here

———

2.2   The model with background points included

We now extend the model by the following assumptions. Conditional on (n, k), x_b is independent of x_c and forms a binomial process on W, i.e. the $m = n - k$ background points are independent and uniformly distributed on W. Further, k conditional on n is binomially distributed with index n and unknown probability $q \in [0, 1]$. So writing $x_c = \emptyset$ (the empty set) if $k = 0$, and $x_b = \emptyset$ if $m = 0$, the density of $x = (x_c, x_b)$ is

$$\pi(x_c, x_b | q, p, \sigma) = \binom{n}{k} q^k \left(\frac{1-q}{|W|} \right)^m \prod_{i=1}^{k} f(x_i | x_1, \ldots, x_{i-1}; p, \sigma) \tag{3}$$

with respect to the measure ν on $\cup_{l=0}^{n} W^l \times W^{n-l}$ given by $\nu = \sum_{l=0}^{n} \nu_k$, where ν_l is the product of Lebesgue measure on W^l and Lebesgue measure on W^{n-l} (interpreting 'Lebesgue measure on W^{0}' as the point measure concentrated at $W^0 = \{\emptyset\}$).

One way of simulating x under (3) is by initially setting $x_c = \emptyset$ and $x_b = \emptyset$, generating independent uniform points $y_1, \ldots, y_n$ in W, and independently of these points, generating independent identically distributed random variables $j_1, \ldots, j_n$, where

$$\mathrm{P}(j_i = 1) = 1 - q, \quad \mathrm{P}(j_i = 2) = q(1-p), \quad \mathrm{P}(j_i = 3) = pq.$$

For $i = 1, \ldots, n$, if $j_i = 1$ then y_i becomes a background point, meaning that y_i is added to x_b; if $j_i = 2$ then y_i becomes an independent cluster point, so it is added to x_c; if $j_i = 3$ then y_i is transformed into a dependent cluster point in accordance to (B)-(D) in Section 2.1, and this dependent cluster point is then added to x_c.

For practical reasons we have ignored possible edge effects caused by unobserved points outside W. Observe that if n is assumed to be Poisson distributed with parameter $\kappa > 0$, then $y = \{y_1, \ldots, y_n\}$ becomes a homogeneous Poisson process, and thus the independent division of y into background points, independent cluster points, and 'points which are later transformed into dependent cluster points' as described above form three independent homogeneous Poisson processes on W given by $\{y_j : i_j = 1\}$, $\{y_j : i_j = 2\}$, $\{y_j : i_j = 3\}$ and with intensities $(1-q)\kappa$, $q(1-p)\kappa$, $pq\kappa$, respectively. To eliminate the problem with edge effects, we could extend this to three independent stationary Poisson processes Φ_1, Φ_2, Φ_3 on $\mathbb{R}^2$, with intensities as above, and where we attach an arrival time to each cluster point in order to specify a sequential construction of the dependent cluster points which are obtained by transforming the points in Φ_3 in exactly the same way as described above. For instance, these arrival times could be independent and identically distributed and follow a continuous distribution (as this ensures that the arrival times are almost surely distinct); see

4

Møller et al. (2010) for a somewhat similar extension in the case of the Matérn hard core process of type III. The resulting point process would be given by the superposition $X = \Phi_1 \cup \Phi_2 \cup \Psi_3$, where Ψ_3 denotes the point process of dependent cluster points, and X is stationary with intensity κ. However, dealing with such an extended model would make it much harder to perform a detailed statistical analysis, so we refrain any further in this paper to consider this extension of the model.

2.3   Data examples and simulations

This section provides further details on the data sets given in Figure 1, and discusses some simulations of our sequential point process model in relation to the data sets.

2.3.1   Location of barrows

Our first data set (Figure 1a) is the location of barrows within a 15×15 km region in Western Jutland, Denmark. Barrows, which are bronze age burial sites resembling small hills, are important sources of information for archaeologists. Contrary to other areas of Denmark, a relatively large proportion of the barrows are still present in the Western Jutland due to less intensive agriculture. The spatial distribution of barrows across Denmark shows a variety of patterns, particularly clusters of points along various lines, where some lines seem to stretch across the landscape for hundreds of kilometers.

The barrow lines have traditionally been regarded as reflecting a prehistoric system of roads, cf. Müller (1904), though there are other potential explanations for this phenomenon, see e.g. Sahlquist (2001). Our sequential point process model provides an alternative explanation, since it has the following naïve interpretation: the barrows are placed according to a 'local decision-making rule', where

- the i'th barrow corresponds to a person who died at a uniformly placed location y_i (independently of previously generated barrows);

- with probability $1 - pq$ the survivors bury the person in a barrow at location y_i, and conditional on this event, with probabilities $(1-q)/(1-pq)$ and $(q-pq)/(1-pq)$, y_i becomes a background point respective an independent cluster point;

- otherwise the survivors bury the person in a barrow close to the closest barrow located at a previous cluster point, and the location of this new barrow becomes a dependent cluster point.

Figure 4 shows a simulated point pattern with the same 15×15 km region and number of points as in the barrows data set, and where $(q, p, \sigma) = (0.758, 0.723, 68.3 \,\mathrm{m})$ is the estimated posterior mean obtained by the Bayesian method in Section 3. Figure 1a (the barrows data set) and Figure 4 (the simulation) have both some similarities and differences. In particular the linear structures in Figure 1a are longer than in Figure 4; this was also typically observed when we made a number of further simulations (not shown here). A closer look at Figure 1a reveals a possible reason for this: the longer linear structures have both small and large gaps between neighboring points, and it is quite possible that the model 'sees' the large gaps as separating different clusters, so the data is 'interpreted' as containing many smaller clusters. This may suggest to replace the 'normal' density function (2) by a density with heavier tails. We shall not pursue this any further in this paper, but merely note that the current model only fits some aspects of the barrow data set.

———

Figure 4 about here

———

2.3.2   Location of mountain tops

Our second data set (Figure 1b) is the location of mountain tops (above $2600 \,\mathrm{m}$ of sea level) in a 7.5×10.5 km region in the Pyrenees of Northern Spain. Many of the mountain tops are located along linear structures, which of course is a consequence of the fact that many tops are located along the mountain ridges. However, visually the linear structures are somewhat obscured by some tops located off the ridges. We consider our

sequential point process model only as a flexible way of modelling these mountain tops and do not intend to give an interpretation in terms of the model's sequential construction.

Figure 3b shows a simulated point pattern with the same $7.5 \times 10.5\,\text{km}$ region and number of points as in the mountain data set, and where $(q, p, \sigma) = (0.825, 0.887, 278.1\,\text{m})$ is the estimated posterior mean obtained by the Bayesian method in Section 3. Comparing with the data in Figure 1b, we see no obvious discrepancies. Both contain medium length linear structures, gaps with no or few points, and quite many solitary points.

3 Simulation-based Bayesian inference

3.1 Likelihood and missing data

The likelihood function for the observed point pattern $\{x_1, \ldots, x_n\}$ can in principle be derived from the density (3) but it turns out to be too complicated because of the following 'missing data' problem. Recall that (3) is a joint density for $x_c = (x_1, \ldots, x_k)$ and $x_b = (x_{k+1}, \ldots, x_n)$. Let $z = (z_1, \ldots, z_n)$ denote an arbitrary ordering of $\{x_1, \ldots, x_n\}$. Define $u = (u_1, \ldots, u_n)$ such that $u_i = 1$ if z_i is a cluster point, and $u_i = 0$ if z_i is a background point. Given the values of $u_1, \ldots, u_n$, define the permutation $\omega = (\omega_1, \ldots, \omega_k)$ of the sequence of those i with $u_i = 1$ such that $x_c = (z_{\omega_1}, \ldots, z_{\omega_k})$, and the permutation $\eta = (\eta_1, \ldots, \eta_m)$ of the sequence of those i with $u_i = 0$ such that $x_b = (z_{\eta_1}, \ldots, z_{\eta_m})$. In other words, z_{ω_i} is the i'th cluster point, and z_{η_j} is the j'th background point. Clearly (x_c, x_b) is in a one-to-one correspondence with (z, u, ω, η), so (u, ω, η) is the missing data, and the density of z is given by summing in (3) over all (u, ω, η). For fixed (u, ω), (3) does not depend on η, so essentially it is only (u, ω) which is missing. Hence the number of terms in the sum reduces to $\sum_{k=0}^{n} n!/k!$ (the cardinality of the state space of (u, ω)), and apparently the sum cannot be reduced any further. This number of terms is very huge when $n = 81$ as in the mountain tops data set, and extremely huge when $n = 1147$ as in the barrows data set. This makes the likelihood based on the data z intractable.

We treat this missing data problem by including (u, ω) into the posterior considered in Section 3.2. Note that conditional on z, (u, ω) is in a one-to-one correspondence with $(x_c, \{x_{k+1}, \ldots, x_n\})$, and its probability mass density is proportional to

$$\tilde{\pi}(u, \omega | z; q, p, \sigma) = \frac{1}{k!} q^k \left(\frac{1-q}{|W|} \right)^m \prod_{i=1}^{k} f(x_i | x_1, \ldots, x_{i-1}; p, \sigma) \tag{4}$$

for all $u = (u_1, \ldots, u_n) \in \{0, 1\}^n$ and all permutations $\omega = (\omega_1, \ldots, \omega_k)$ of the sequence of those i with $u_i = 1$. As discussed in Section 3.2, as one ingredient of making simulation-based Bayesian inference, we make simulations from (4). Incidentally, such simulations would also be needed if we want to determine an approximate maximum likelihood estimate of θ based on a missing data Markov chain Monte Carlo (MCMC) approach, see Geyer (1999) and Møller and Waagepetersen (2004).

3.2 Prior assumptions and posterior distribution

As argued in Section 3.1, we shall consider the posterior distribution of the unknown quantities u, ω, q, p, σ. By Bayes formula, the posterior density is proportional to the term (4) times the prior density of (q, p, σ). In the sequel we impose independent and 'flat' priors on the parameters q, p, and σ, where both p and q follow a uniform distribution on the interval $[0, 1]$, while σ follows an inverse gamma density $\pi(\sigma)$ with shape parameter 2 and a known scale parameter $\beta > 0$. Thereby σ has mean value β but an infinite variance. These prior assumptions, including the specification of β, seem less important as the posterior distribution based on either of our two data sets will be dominated by the term (4) as demonstrated in Section 3.3.

The posterior distribution of $(u, \omega, q, p, \sigma)$ given z naturally splits into four 'full conditionals', where we use the generic notation $\tilde{\pi}$ for unnormalized densities:

(I) $(u, \omega) | (z, q, p, \sigma)$ has unnormalized probability mass density $\tilde{\pi}(u, \omega | z, p, q, \sigma)$ given by (4);

(II) $q | (z, u, \omega, p, \sigma)$ is Beta-distributed with parameters $k + 1$ and $m + 1$;

(III) $p|(z, u, \omega, q, \sigma)$ has unnormalized density

$$\tilde{\pi}(p|z, u, \omega, q, \lambda) = \prod_{i=1}^{k} f(z_{\omega_i}|z_{\omega_1}, \ldots, z_{\omega_{i-1}}; p, \sigma), \quad 0 < p < 1; \tag{5}$$

(IV) $\sigma|(z, u, \omega, q, p)$ has unnormalized density

$$\tilde{\pi}(\sigma|z, u, \omega, q, p) = \pi(\sigma) \prod_{i=1}^{k} f(z_{\omega_i}|z_{\omega_1}, \ldots, z_{\omega_{i-1}}; p, \sigma), \quad \sigma > 0. \tag{6}$$

We presume that the reader has some familiarity to MCMC algorithms, see e.g. Gilks, Richardson and Spiegelhalter (1996). For simulation of the posterior, we use a fixed scan Metropolis-within-Gibbs algorithm (also called a hybrid Metropolis-Hastings algorithm) which makes updates from the full conditionals in (I)-(IV). Details are given in Appendix A.

3.3 Posterior results

In this section we report on simulation-based posterior results when $\beta = 150$; results for other choices of β look very similar as discussed in Section 3.4. For generating simulations of the posterior distributions for the mountain tops data set, we have used 100,000 steps in the Markov chains, where we have discarded the first 10,000 steps as burn-in. Since the barrows data set is much larger than the mountain tops data set, and it thus takes more steps to adequately explore the space of missing data, we have used 500,000 steps with a burn-in of 100,000 steps for this data set. For approximating the posterior distributions of the parameters (q, p, σ), we have used all simulations resulting from this, while for approximating the posterior distributions of the missing data (u, ω), we have subsampled the data, only using every 100th sample for the approximations.

For our two data sets, Figures 5 and 6 show the marginal posterior distributions of the parameters q, p, and σ, and Figures 7 and 8 the bivariate posterior distributions of (q, p), (q, σ), and (p, σ). These distributions are very different from the flat and independent priors assumed in Section 3.2, showing that the posterior density is dominated by the term (4) and not by the prior density of (q, p, σ).

———
Figure 5 about here
———

———
Figure 6 about here
———

———
Figure 7 about here
———

———
Figure 8 about here
———

For the barrows data set, the marginal and posterior distributions of the parameters q, p, σ (Figure 6) and the bivariate posterior distributions of (q, p), (q, σ), and (p, σ) (Figure 8) are rather close to normal and bivariate normal distributions. There is a positive correlation between σ and q, and between σ and p, while q and p are negatively correlated—this may seem logical for the following reason. As σ increases, the 'attraction' of a dependent cluster point towards its closest previous cluster point is relaxed, so both the mean number of cluster points (i.e. nq) and the probability of a cluster point being a dependent cluster point (i.e. p) tend to increase but in a way such that mean number of dependent cluster points (i.e. npq) is

7

not getting too large. For the mountain tops data set, it looks quite differently, as the marginal posterior distributions of q and p in Figure 5 are left skewed, and the bivariate distributions in Figure 7 are far from being normal.

Figure 9 illustrates the usefulness of our Bayesian approach for estimating the linear structures. Each data point in the figure is the center of a circle with a radius proportional to the marginal posterior probability of the data point being a cluster point, where in each figure the largest circle corresponds to a probability of one. Figure 9a concerns the mountain tops data set, where the type of the points seem to be well-estimated, since all the points located in the linear formations are marked by large circles, and the solitary points are marked by small circles. In Figure 9b (the barrows data set) this is mostly also the case, although some points may seem to have been 'misclassified'.

———

Figure 9 about here

———

Figure 10 shows the estimated order in which the cluster points appear in linear structures using circles of different radii. More precisely, consider the order $(\omega_1, \ldots, \omega_k)$ of the cluster points, and let o_i denote the number of the point z_i in this order, i.e. $o_i = j$ if $\omega_j = i$, and let $\bar{o}_i$ denote the average of a sample of o_i obtained from the Markov chains, disregarding all steps where z_i is a background point. The figures show circles centered at the data points with radii proportional to $\exp(-b\bar{o}_i)$ where $b = 0.004$ for the barrows data set and $b = 0.1$ for the mountain tops data set, i.e. large circles in the figures indicate that points have appeared early. The b values have been chosen by trial-and-error to give plots where the different sizes of the circles can be clearly seen. Note that good choices of b depends on the number of points n and the configuration of the points. In Figure 10a we see that the early points are located in the largest cluster, while the solitary points are usually appearing very late, when they are part of the cluster (they are usually background points as can be seen by comparing Figure 10a with Figure 9a). Figure 10b shows a different picture: here the larger clusters can be seen as consisting of many smaller clusters of points appearing roughly at the same time. This indicates that these are indeed interpreted as many smaller clusters as suggested in Section 2.3.

———

Figure 10 about here

———

3.4   Model checking and sensitivity analysis

In this section, we check how well the model fits certain aspects of the data, as well as how sensitive the model is to changes in the prior distribution.

3.4.1   Model checking

We have already compared the two data sets to simulations visually in Section 2.3. In the present section, we compare simulated data from the posterior predictive distribution with the observed data; recall that a simulation from the posterior predictive distribution is given by simulating first a realization (p, q, σ) from the posterior distribution and second a realization of a point pattern generated by our sequential point process model with parameters p, q, σ. We focus on investigating the shape of the linear structures on a more detailed level than in Section 2.3.

One way of measuring the shape of the linear structures is to use the angle between an arbitrary point and the two nearest points. Consider the point z_i, and let z_j and z_k denote its two nearest points. Then we denote the angle between the vectors $z_j - z_i$ and $z_k - z_i$ by $a_i \in [0, \pi]$. An abundance of angles close to either 0 or π indicates that many points are located on roughly straight lines. In Figure 11, the crosses correspond to a histogram of such angles for the data grouped into ten equidistant intervals of $[0, \pi]$, and the small horizontal lines correspond to 0.5 %, 2.5 %, 50 %, 97.5 %, and 99.5 % quantiles obtained from 199 posterior predictions. Figure 11a shows no clear discrepancies between the mountain tops data set and the simulations, although for two values the data lies on the edge of approximated 99% interval. In Figure 11b

there some more clear differences, in particular the simulations show a higher tendency to produce angles close to 0 than the barrows data set.

———

Figure 11 about here

———

Another kind of model check is based on 'squeezedness' defined as follows. Consider the Delaunay triangulation generated by the data, i.e. if z_i and z_j share a common Dirichlet edge, then they are the end points of an edge in the Delaunay triangulation (see e.g. Okabe et al. (1992) or Møller (1994)). Consider two Delaunay triangles with vertices $\{z_i, z_j, z_k\}$ and $\{z_i, z_j, z_l\}$, i.e. z_i and z_j specify a common edge $e_{i,j}$ of the two triangles. Then

$$q_{i,j} = 1 - \|z_i - z_j\| / \min\{(\|z_i - z_l\| + \|z_j - z_l\|)/2, (\|z_i - z_k\| + \|z_j - z_k\|)/2\}$$

is the squeezedness associated to $e_{i,j}$. This quantity takes values in the interval $[-1, 1]$, and it is close to one if the points z_i and z_j are clearly part of a linear structure with z_k and z_l on each side of this structure. Figure 12a shows the empirical distribution function of $q_{i,j}$ for the mountain tops data set together with quantiles obtained from 199 posterior predictions. The empirical distribution function for the data resembles the simulations fairly well and stays within the quantiles, except for a brief excursion around the value $q_{i,j} = 0.2$. Here the empirical distribution function is lower than the simulations, suggesting that the linear structures in the data is slightly tighter packed than in the simulations. Figure 12b shows a similar picture for the barrows data set, except here the empirical distribution function for the data is higher than the simulations around the value $q_{i,j} = 0.5$, suggesting more loosely packed linear structures in the data. Note that the high end of these figures, i.e. values where $q_{i,j}$ is close to one, is the most interesting part of the figures, since points in linear structures typically will influence this part of the function.

———

Figure 12 about here

———

3.4.2 Sensitivity analysis

In this section, we check how sensitive the posterior distributions are to changes in the values of the hyper-parameters. We focus on β since this parameter has been chosen rather arbitrarily to be 150 in Section 3.3. For the mountain tops data set we have also tried other values to test whether this choice has any significant influence on the posterior distributions. Here we compare the values $\beta = 50, 150, 300$. The two other values, $\beta = 50$ and $\beta = 300$, both give posterior distributions for the parameters that are visually almost indistinguishable from the ones shown in Figures 5 and 7. The posterior distributions of the missing data for the different choices of β are even closer; for all three values of β the plots of the posterior distributions of the missing data are visually the same as the ones given in Figures 9a and 10a.

Since the barrows data set contains many more points than the mountain tops data set, there is much more information in the data, and we therefore expect the prior distribution to have even less impact on the analysis for this data set.

As a further argument that the present choice of prior distributions does not influence the analysis much, consider the estimated marginal posteriors in Figures 5 and 6. All of these look very different from the uniform and inverse gamma distributions used as priors, which indicates that the data must have a high impact on the posterior distributions.

Appendix A: Posterior simulations

This appendix describes how to make MCMC updates from each of the full conditionals (I)-(IV) in Section 3.2.

To update q from the Beta-distribution in (II), we simply use a Gibbs update, while we use random walk Metropolis algorithms in (III)-(IV). Specifically, in case of (III), if p is the current state of the random walk

Metropolis chain, we generate a proposal p' from the uniform distribution on $[p - \epsilon, p + \epsilon]$ for some fixed value $\epsilon > 0$, calculate the Hastings ratio $H(p, p') = \mathbf{1}[0 < p' < 1]\tilde{\pi}(p'|z, u, \omega, q, \sigma)/\tilde{\pi}(p|z, u, \omega, q, \sigma)$ obtained from (5), accept the proposal with probability $\min\{1, H(p, p')\}$ as the next state of the chain, and otherwise let p be the next state of the chain. Similarly in case of (IV), where if σ is the current state of the chain, we generate a normally distributed proposal σ' with mean σ and some fixed standard deviation τ, calculate the Hastings ratio $H(\sigma, \sigma') = \mathbf{1}[\sigma' > 0]\tilde{\pi}(\sigma'|z, u, \omega, q, p)/\tilde{\pi}(\sigma|z, u, \omega, q, p)$ obtained from (6), and make a similar accept/rejection step to see if σ or σ' should be the next state.

For (I) we use a Metropolis-within-Gibbs algorithm, with updates of different types as described in the sequel. Let (u, w) denote a generic 'current state of the chain', and let x_c and x_b be the corresponding 'current states' of the cluster and background processes. For $i = 1, \dots, n$, let x_{k+1} be the first z_i with $u_i = 0$, x_{k+2} the second z_i with $u_i = 0$, and so on. This assignment plays no role as the ordering of background points is of no importance, cf. Section 3.2. In Sections A1, A2, and A3, we describe updates of type 0, 1, and $i \geq 2$, respectively. Briefly, in a type 0 update, we propose that a background point becomes a cluster point, in a type 1 update, we propose that a cluster point becomes a background point, and in a type $i \geq 2$ update, we propose to shift the ordering of two succeeding cluster points. Since (4) is a discrete density, it becomes straightforward to calculate the Hastings ratios. We assume there is an equal probability for doing either a type 0 or type 1 update, and we use a systematic updating scheme, where we first do either a type 0 or type 1 update, and second if $k > 1$, make updates of type $2, \dots, k$.

In the analyses in Section 3.3, we have used $\epsilon = 0.1$ and $\tau = 10$ when running the Markov chains. For the mountain tops data set 100,000 steps with a burn-in of 10,000 steps was used, while the barrows data set required 500,000 steps with a burn-in of 100,000. The reason for the higher number of steps and the high burn-in required for the barrows data set is the higher number of points in the data set, which means that we need many type 0 and 1 updates for adequately exploring the space of the missing data. Apart from this, the mixing of the Markov chains seems to be good. Note also that the computation time for each step increases with the number of cluster points.

A1: A background point becomes a cluster point

For a type 0 update, if $k = n$ we do nothing, so suppose that $m = n - k > 0$. Uniformly at random we pick $j \in \{0, \dots, k\}$ and $i \in \{k+1, \dots, n\}$, and we propose to update k by $k' = k + 1$, m by $m' = m - 1$, x_c by

$$x_c' = (x_1', \dots, x_{k'}') = (x_1, \dots, x_j, x_i, x_{j+1}, \dots, x_k)$$

and x_b by

$$x_b' = (x_{k'+1}', \dots, x_n') = (x_{k+1}, \dots, x_{i-1}, x_{i+1}, \dots, x_n).$$

In other words, we propose to update (u, ω) by

$$u' = (u_1', \dots, u_n') = (u_1, \dots, u_{i-1}, 1, u_{i+1}, \dots, u_n)$$

and

$$\omega' = (\omega_1', \dots, \omega_{k+1}') = (\omega_1, \dots, \omega_j, \omega_{j+1}', \omega_{j+1}, \dots, \omega_k)$$

where ω_{j+1}' is defined by $x_i = z_{\omega_{j+1}'}$.

To obtain the Hastings ratio, we start by recalling that (u, ω) is a one-to-one correspondence with $(x_c, \{x_{k+1}, \dots, x_n\})$. When a background point becomes a cluster point, there are m possibilities for choosing a background point and $k + 1$ possibilities for choosing a position in the order of the clusters points. For the reverse update, i.e. when a cluster point becomes a background point (the type 1 update in Section A2), there are k' possibilities for choosing a cluster point, and since the order of the background points is irrelevant, we do not have to account for the position among the background points. From this and (4) we obtain the Hastings ratio $H_0 = r(j, x_c, x_i)$, where

$$r(j, x_c, x_i) = \frac{\frac{1}{k'} \frac{1}{k'!} q^{k'} \left(\frac{1-q}{|W|}\right)^{m'} \prod_{l \leq k'} f(x_l'|x_1', \dots, x_{l-1}')}{\frac{1}{k+1} \frac{1}{m} \frac{1}{k!} q^k \left(\frac{1-q}{|W|}\right)^m \prod_{l \leq k} f(x_l|x_1, \dots, x_{l-1})}.$$

Note that we suppress in the notation that $f(x_l|x_1,\ldots,x_{l-1}) = f(x_l|x_1,\ldots,x_{l-1};p,\sigma)$ and $f(x'_l|x'_1,\ldots,x'_{l-1}) = f(x'_l|x'_1,\ldots,x'_{l-1};p,\sigma)$ both depend on (p,σ). Clearly, the Hastings ratio reduces to

$$r(j,x_c,x_i) = \frac{mq|W|}{(k+1)(1-q)} f(x_i|x_1,\ldots,x_j) \prod_{l=j+1}^{k} \frac{f(x_l|x_1,\ldots,x_{l-1},x_i)}{f(x_l|x_1,\ldots,x_{l-1})}.$$

A2: A cluster point becomes a background point

Type 1 updates are the reverse of type 0 updates. If $k=0$ we do nothing, so suppose that $k>0$. Uniformly at random we pick $j \in \{1,\ldots,k\}$ and $i \in \{k,\ldots,n\}$, and propose $k'=k-1$, $m'=m+1$,

$$x'_c = (x'_1,\ldots,x'_{k'}) = (x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_k)$$

and

$$x'_b = (x'_{k'+1},\ldots,x'_n) = (x_{k+1},\ldots,x_i,x_j,x_{i+1},\ldots,x_n).$$

In other words, we propose

$$u' = (u'_1,\ldots,u'_n) = (u_1,\ldots,u_{i-1},0,u_i,\ldots,u_n)$$

and

$$\omega' = (\omega'_1,\ldots,\omega'_{k-1}) = (\omega_1,\ldots,\omega_{j-1},\omega_{j+1},\ldots,\omega_k).$$

The Hastings ratio is $H_1 = 1/r(j-1,x'_c,x_j)$.

A3: Shifting the ordering of two succeeding cluster points

When $k>1$ and $i \in \{2,\ldots,k\}$, a type i update is given as follows. Propose to keep $x'_b = x_b$ and replace x_c by

$$x'_c = (x'_1,\ldots,x'_k) = (x_1,\ldots,x_{i-2},x_i,x_{i-1},x_{i+1},\ldots,x_k).$$

In other words, propose $u'=u$ and

$$\omega' = (\omega'_1,\ldots,\omega'_k) = (\omega_1,\ldots,\omega_{i-2},\omega_i,\omega_{i-1},\omega_{i+1},\ldots,\omega_k).$$

By (4), the Hastings ratio is

$$H_i = \frac{f(x_i|x_1,\ldots,x_{i-2})f(x_{i-1}|x_1,\ldots,x_{i-2},x_i)}{f(x_{i-1}|x_1,\ldots,x_{i-2})f(x_i|x_1,\ldots,x_{i-1})}.$$

References

Berman, M. (1986). Testing for spatial association between a point pattern and another stochastic process, *Applied Statistics* **35**: 54–62.

Berman, M. and Diggle, P. (1989). Testing for spatial association between a point pattern and another stochastic process, *Journal of Royal Statistical Society Series B* **51**: 81–92.

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

Blackwell, P. G. and Møller, J. (2003). Bayesian analysis of deformed tessellation models, *Advances in Applied Probability* **35**: 4–26.

Foxall, R. and Baddeley, A. (2002). Nonparametric measures of association between a spatial point process and a random set, with geological applications, *Applied Statistics* **51**: 165–182.

Geyer, C. J. (1999). 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*, Chapman & Hall/CRC, Boca Raton, Florida, pp. 79–140.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London.

Møller, J. (1994). *Lectures on Random Voronoi Tessellations*, Lecture Notes in Statistics 87, Springer-Verlag, New York.

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

Møller, J. and Waagepetersen, R. P. (2007). Modern spatial point process modelling and inference (with discussion), *Scandinavian Journal of Statistics* **34**: 643–711.

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

Müller, S. (1904). Vei og bygd i sten- og bronzealderen, *Aarbøger for Nordisk Oldkyndighed og Historie* pp. 1–64.

Okabe, A., Boots, B. and Sugihara, K. (1992). *Spatial Tessellations. Concepts and Applications of Voronoi Diagrams*, Wiley, Chichester.

Sahlquist, L. (2001). Territorial behaviour and communication in a ritual landscape, *Geografiska Annaler* **83 B**(2): 79–102.

Skare, Ø., Møller, J. and Jensen, E. B. V. (2007). Bayesian analysis of spatial point processes in the neighbourhood of Voronoi networks, *Statistics and Computing* **17**: 369–379.
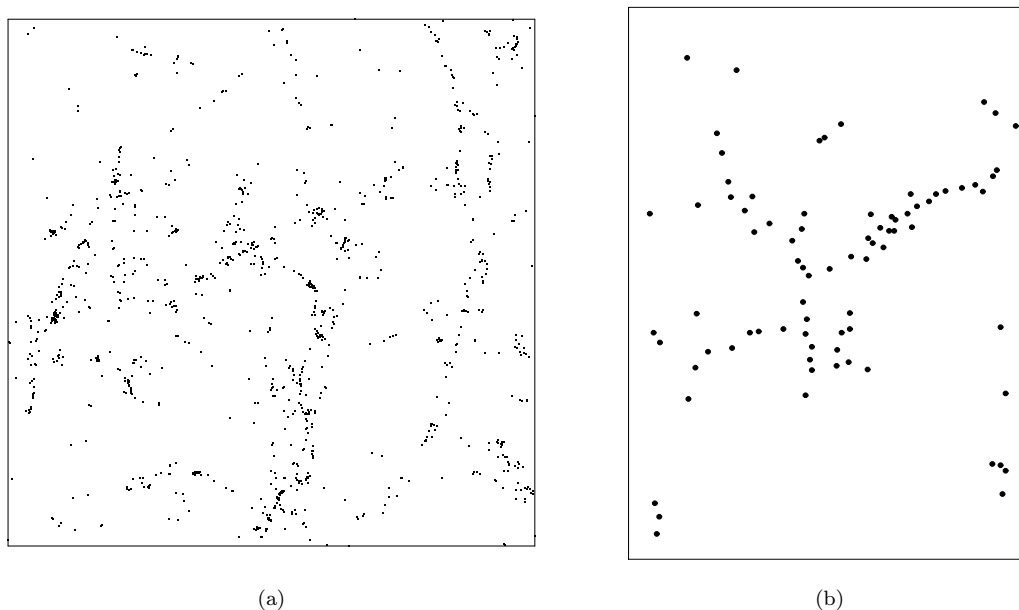
Figure 1: (a): The locations of 1147 barrows (bronze age burial sites) in a 15×15 km region. (b): The locations of 81 mountain tops in a 7.5×10.5 km region.
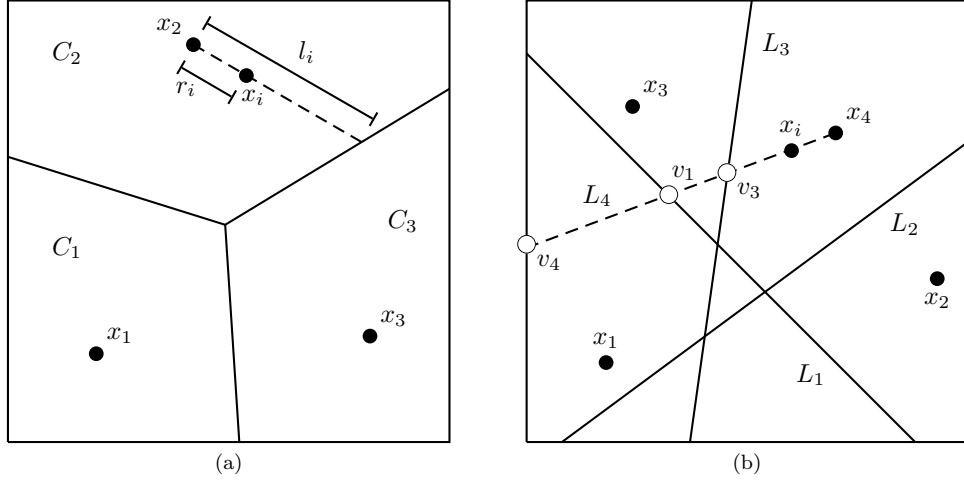
Figure 2: (a): Example with a rectangular window W, $i = 4$, three cluster points x_1, x_2, x_3, and their respective Dirichlet cells C_1, C_2, C_3 within W. The new cluster point x_i and the distances l_i and r_i are shown. (b): Example with $i = 5$, showing four cluster points $x_1, \ldots, x_4$ and a new cluster point x_i (filled circles), where x_i is closest to x_4. The half line L_4 (dashed line), the lines L_1, L_2, L_3 (solid lines), and the intersection points v_1, v_3 and v_4 (empty circles) are also shown.
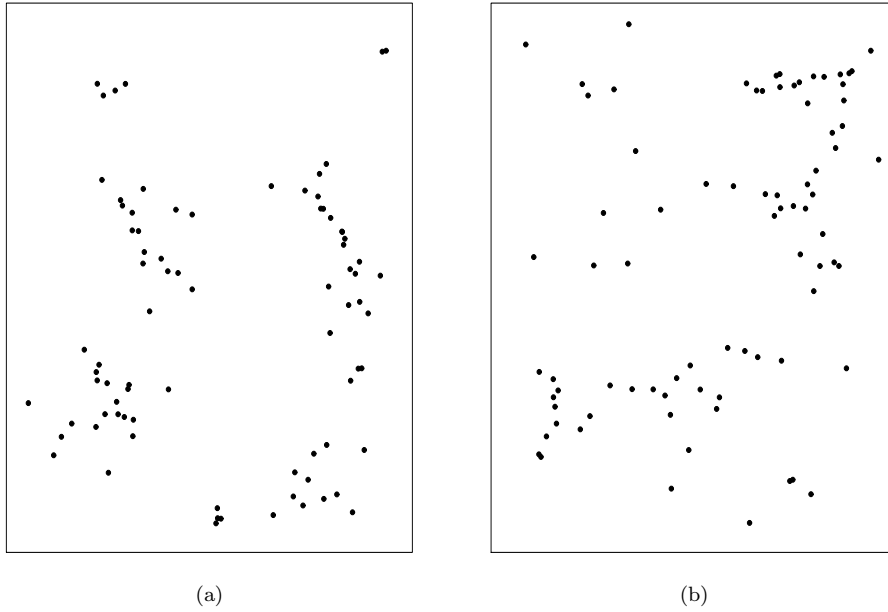
Figure 3: (a): A simulated point pattern of $k = 81$ cluster points when $p = 0.800$, $\sigma = 286.4\,\text{m}$, and W is a $7.5 \times 10.5\,\text{km}$ region. (b): A simulated point pattern of $n = 81$ cluster and background points, where $q = 0.825$, $p = 0.887$, $\sigma = 278.1\,\text{m}$, and W is a $7.5 \times 10.5\,\text{km}$ region.

13
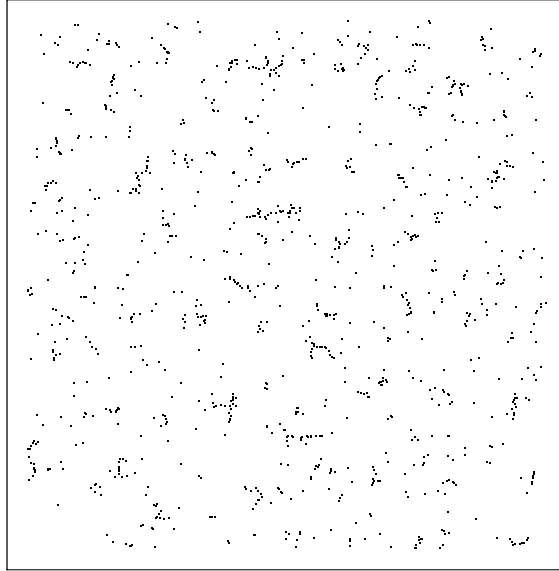
Figure 4: A simulated point pattern of $n = 1147$ cluster and background points, where $q = 0.758$, $p = 0.723$, $\sigma = 68.3\,\text{m}$, and W is a $15 \times 15\,\text{km}$ region.
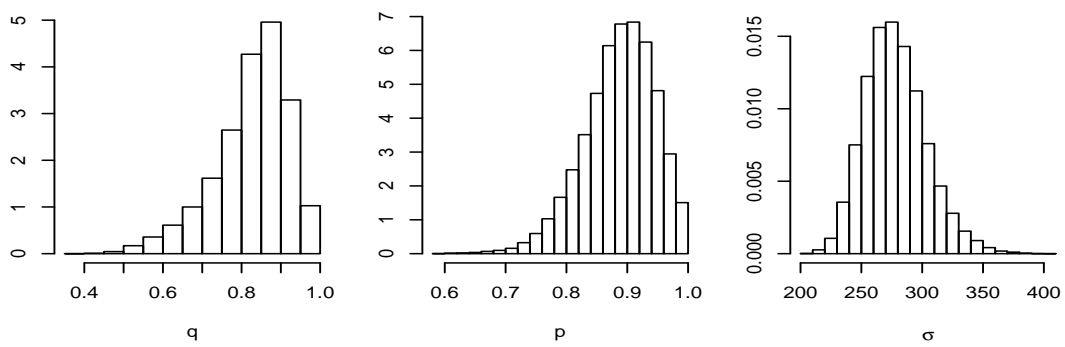
Figure 5: Histograms showing the marginal posterior distributions of q, p, and σ for the mountain tops data set.
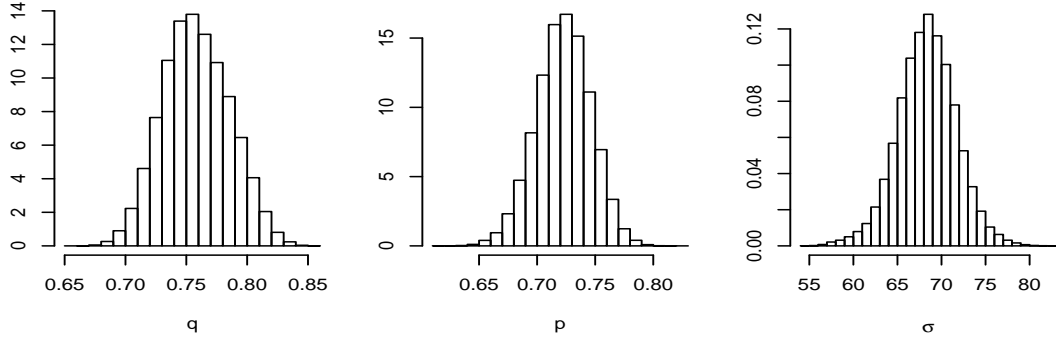
Figure 6: Histograms showing the marginal posterior distributions of q, p, and σ for the barrows data set
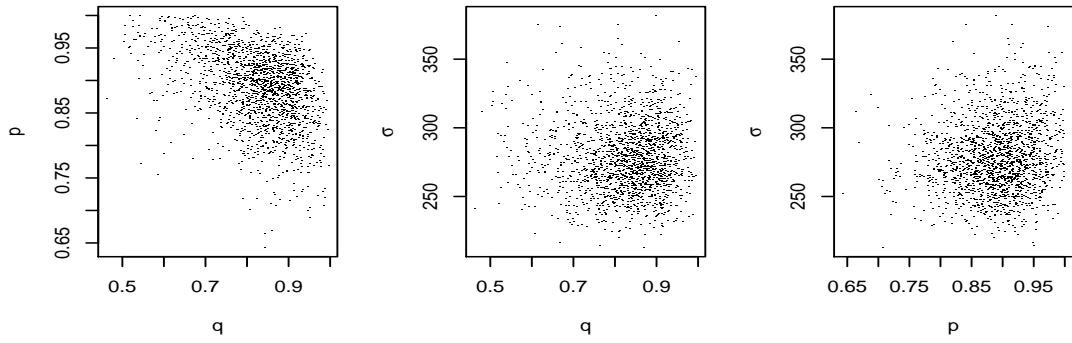

Figure 7: Scatter plots showing the marginal posterior distributions of (q, p), (q, σ), and (p, σ) for the mountain tops data set. In each plot the dots are obtained by subsampling from an MCMC sample obtained by taking every 50th sample.
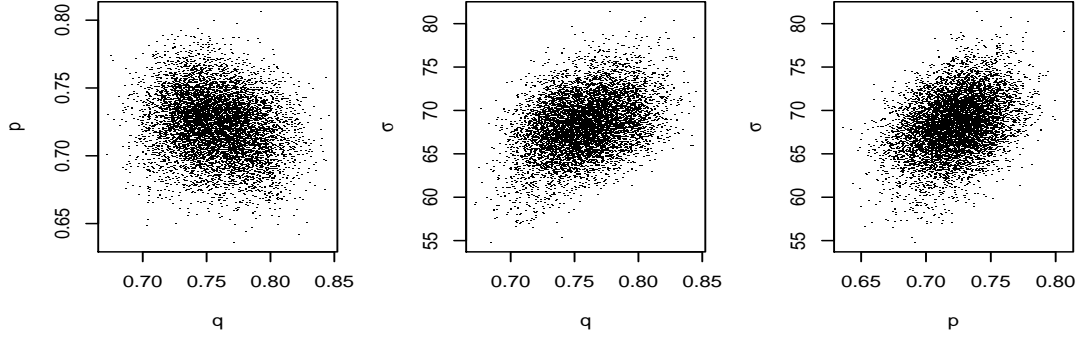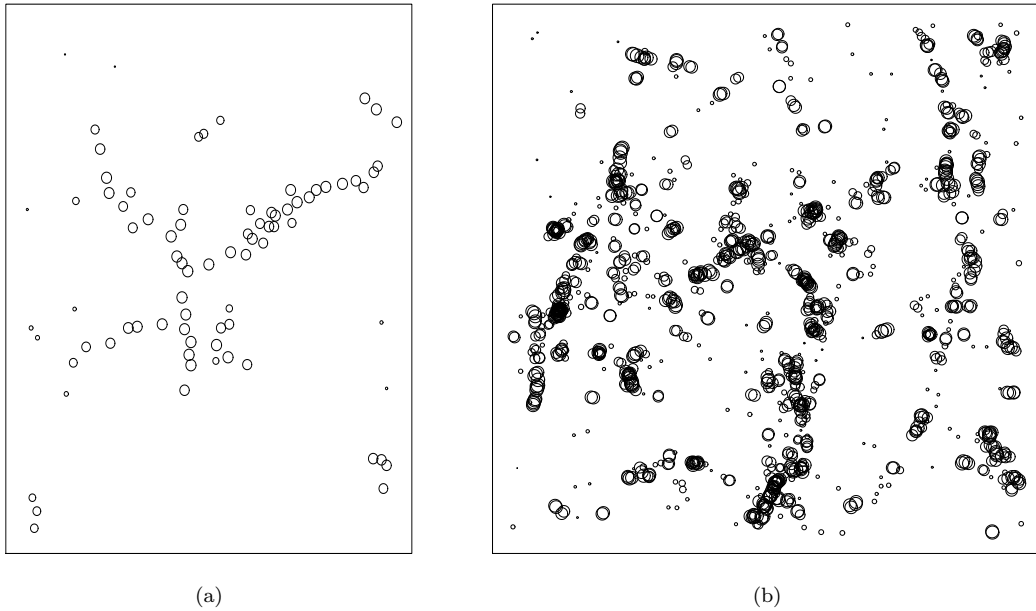
Figure 8: Scatter plots showing the marginal posterior distributions of (q, p), (q, σ), and (p, σ) for the barrows data set. In each plot the dots are obtained by subsampling from an MCMC sample obtained by taking every 50th sample.

Figure 9: Scatter plot with circles indicating the marginal posterior probability of a point being a cluster point for (a) the mountain tops data set and (b) the barrows data set.
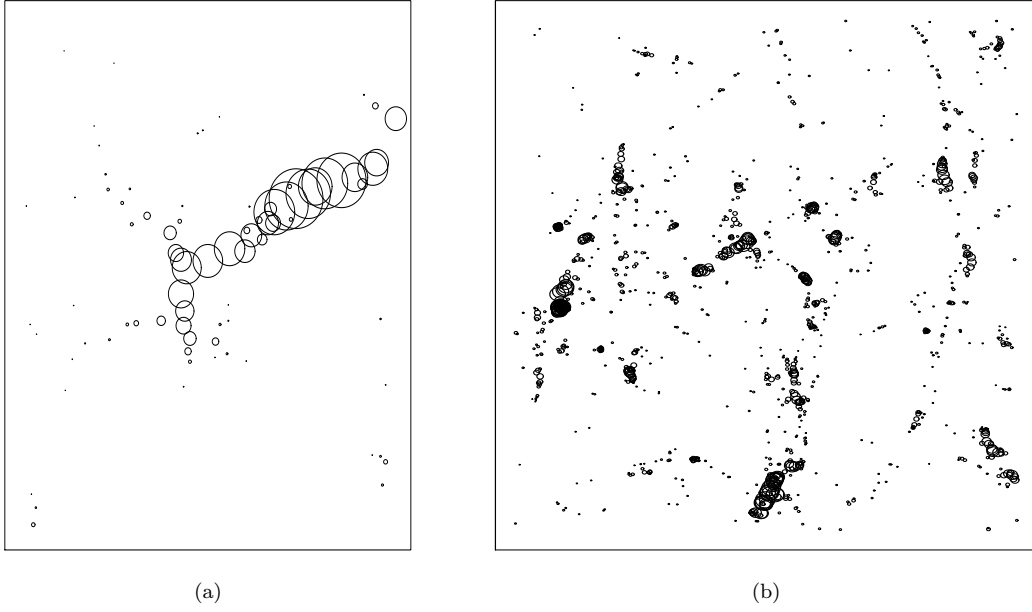
Figure 10: Scatter plot with circles indicating the order in which the cluster points occur in (a) the mountain tops data set and (b) the barrows data set.
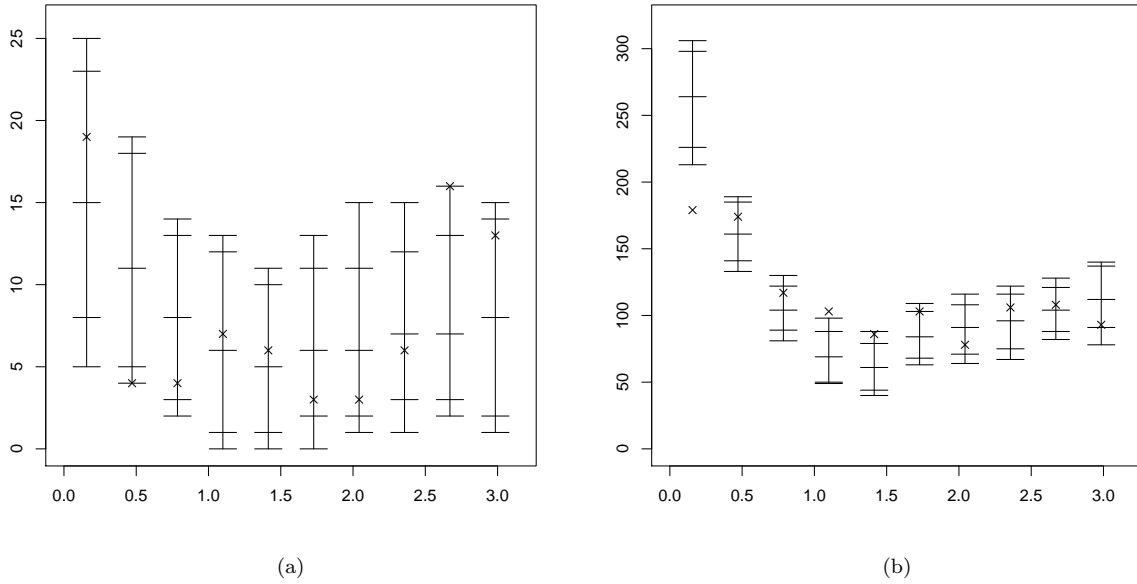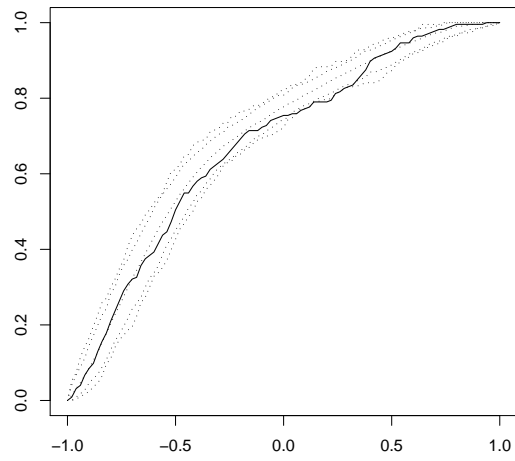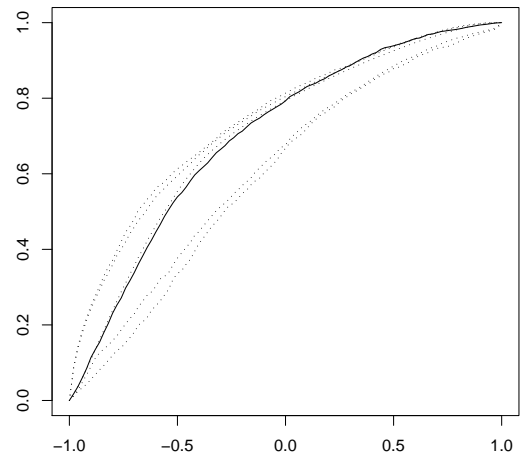

Figure 11: (a): The distribution of angles a_i for the mountain tops data set (crosses) and 0.5 %, 2.5 %, 50 %, 97.5 %, and 99.5 % quantiles obtained from 199 posterior predictions. (b): A similar plot for the barrows data set.

17

(a)   (b)

Figure 12: (a): The distribution function for squeezedness $q_{i,j}$ for the mountain tops data set (solid line) and 0.5 %, 2.5 %, 50 %, 97.5 %, and 99.5 % quantiles obtained from 199 posterior predictions. (b): A similar plot for the barrows data set.

18