

AALBORG UNIVERSITY

**Spatial cluster point processes related to
Poisson-Voronoi tessellations**

by

Jesper Møller and Jakob Gulddahl Rasmussen

R-2013-10

December 2013

DEPARTMENT OF MATHEMATICAL SCIENCES
AALBORG UNIVERSITY

Fredrik Bajers Vej 7 G ■ DK-9220 Aalborg Øst ■ Denmark

Phone: +45 99 40 99 40 ■ Telefax: +45 99 40 35 48

URL: <http://www.math.aau.dk>



Spatial cluster point processes related to Poisson-Voronoi tessellations

Jesper Møller · Jakob Gulddahl Rasmussen

Abstract We discuss how to construct models for cluster point processes within ‘territories’ modelled by d -dimensional Voronoi cells whose nuclei are generated by a latent Poisson process (where the planar case $d = 2$ is of our primary interest). Conditional on the territories/cells, the clusters are independent Poisson processes whose points may be aggregated around or away from the nuclei and along or away from the boundaries of the cells. Observing the superposition of clusters within a bounded region, we discuss how to account for edge effects. Bayesian inference for a particular flexible model is discussed in connection to a botanical example.

Keywords Bayesian analysis · clustering · Cox process · edge effect · intensity estimation · latent point process

1 Introduction

This paper considers a latent stationary point process \mathbf{Y} , the *primary point process*, on \mathbb{R}^d with intensity $\kappa > 0$, and its associated Voronoi (or Dirichlet) tessellation with cells

$$C_i = C(y_i; \mathbf{Y}) = \{x \in \mathbb{R}^d : \|x - y_i\| \leq \|x - y_j\| \text{ for all } y_j \in \mathbf{Y}, j \neq i\}, \quad (1)$$

where $\|\cdot\|$ denotes Euclidean distance. Thus, calling the points of \mathbf{Y} for nuclei, we can think of C_i as a territory consisting of all points in \mathbb{R}^d which are closer to

the nucleus y_i than to any other nucleus y_j ; for background material on Poisson-Voronoi tessellations, see Møller (1989, 1994) and Okabe et al. (2000). In our application example discussed in Section 4, the planar case $d = 2$ is considered.

Conditional on \mathbf{Y} , the *secondary point process* \mathbf{X} is a Poisson process on \mathbb{R}^d with intensity function

$$\Lambda(x) = \alpha + \beta \sum_i 1[x \in C_i] k(A_i) h(x - y_i; C_i - y_i) \quad (2)$$

for $x \in \mathbb{R}^d$. Here $\alpha \geq 0$ and $\beta \geq 0$ are parameters; the sum is over all points in \mathbf{Y} ; $1[\cdot]$ denotes the indicator function; $A_i = |C_i|$ is the Lebesgue measure of C_i (i.e. area if $d = 2$); k is a non-negative (Borel) function; for any bounded convex polytope $C \subset \mathbb{R}^d$ containing the origin o and with $0 < |C| < \infty$, $h(\cdot; C)$ is a density on C with respect to Lebesgue measure; and $C_i - y_i$ denotes C_i translated with $-y_i$. We assume \mathbf{X} is observed within a bounded region $W \subset \mathbb{R}^d$ with $|W| > 0$. Thus we should account for edge effect as discussed in Section 2. Furthermore, Section 3 specifies models for the *offspring density* h .

By (2), \mathbf{X} is stationary, i.e. its distribution is invariant under shifts in \mathbb{R}^d , and \mathbf{X} is a Cox process (Cox 1955) which can be viewed as a union of point processes $\bigcup_i \mathbf{X}_i$ where \mathbf{X}_i is a ‘cluster’ associated to $y_i \in \mathbf{Y}$. Specifically, conditional on \mathbf{Y} , the \mathbf{X}_i ($y_i \in \mathbf{Y}$) are independent Poisson processes defined on the respective cells C_i ($y_i \in \mathbf{Y}$), and the intensity of \mathbf{X}_i at location $x \in C_i$ is $\alpha + \beta k(A_i) h(x - y_i; C_i - y_i)$. Moreover, the cluster \mathbf{X}_i can be viewed as the union of two independent Poisson processes $\mathbf{X}_{i,1}$ and $\mathbf{X}_{i,2}$ on C_i , where $\mathbf{X}_{i,1}$ is homogeneous with intensity α and $\mathbf{X}_{i,2}$ has intensity function $\beta k(A_i) h(\cdot - y_i; C_i - y_i)$. We refer to $\bigcup_i \mathbf{X}_{i,1}$ as the *background process* and $\bigcup_i \mathbf{X}_{i,2}$ as the *cluster process*. For instance, if k is the identity mapping, the mean number of points in $\mathbf{X}_{i,2}$ is βA_i .

Department of Mathematical Sciences, Aalborg University, Fredrik Bajersvej 7E, DK-9220 Aalborg
Tel.: +45-99408863
Fax: +45-99403548
E-mail: jm@math.aau.dk

In the special case where $\beta = 0$, \mathbf{X} is simply a stationary Poisson process independent of \mathbf{Y} , and in relation to telecommunication networks Foss and Zuyev (1996) studied distributional properties of the cluster for a typical cell. If $\alpha = 0$, $\beta = 1$, $k(\cdot) = 1$, and $h(\cdot; C) = 1/|C|$ is uniform on C , then $\Lambda(x) = \sum_i 1[x \in C_i]/A_i$ is an infinite version of Voronoi estimator used for non-parametric intensity estimation (see Barr and Schoenberg (2010) and the references therein). Our main interest is on the case where neither $\beta = 0$ nor $h(\cdot; C)$ is uniform on C , i.e. when each cluster is an inhomogeneous Poisson process on its corresponding cell, and we are in particular interested in detecting the underlying Voronoi tessellation and the offspring density when assuming k is the identity function. We show how a Bayesian approach is useful in this respect. Bayesian approaches closely related to ours include Heikkinen and Arjas (1998), Blackwell (2001), Byers and Raftery (2002), Blackwell and Møller (2003), and Skare et al. (2007).

Section 4 considers a botanical example: Fig. 1 shows the position of 72 daisies on a rectangular observation window of side length 1.6×1.5 (the unit of length used in this dataset is approximately 47.5 cm). A visual inspection of this figure reveals that the daisies form a clustered point pattern. This is confirmed by Fig. 2 which shows the estimated L -function minus the identity; for a clustered point pattern, this summary statistic is expected to be above the zero-line (see e.g. Møller and Waagepetersen(2004)). One explanation for the clustering may be that daisies spread by sending out rhizomes (horizontal underground root-stalks sending out new shoots) resulting in daisies located close together. If we think of the area covered by such a root system as a territory, this can be represented by a Voronoi cell in our model.

The statistical analysis in this paper have been conducted with R. We have developed our own software for the analysis, where we have used many of the functions from the `spatstat` package (Baddeley and Turner, 2005), for example when creating plots of the so-called L , F , G , J and inhomogeneous K functions.

2 Accounting for edge effects

For small point pattern datasets such as in Fig. 1, accounting for edge effects is crucial. This section discusses strategies when observing a Voronoi tessellation within a bounded region $W \subset \mathbb{R}^d$ and generated by a stationary Poisson process \mathbf{Y} which is approximated by a finite subprocess $\tilde{\mathbf{Y}}$.

We let $W \subseteq W_{\text{ext}} \subset \mathbb{R}^d$ be bounded Borel sets, define $\tilde{\mathbf{Y}} = \mathbf{Y} \cap W_{\text{ext}}$, and consider the Voronoi cells \tilde{C}_i

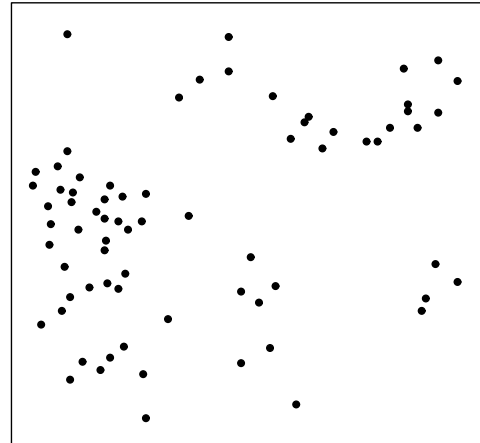


Fig. 1: Daisies dataset: The locations of 72 daisies observed on a 1.6×1.5 observation window.

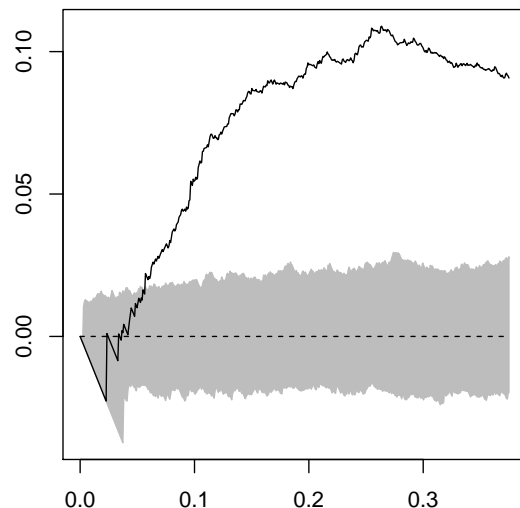


Fig. 2: The estimated L -function (minus the identity) for the Daisies dataset. The grey area is 95% pointwise bounds calculated from 199 simulations of a homogeneous Poisson process.

generated by $\tilde{\mathbf{Y}}$, i.e. when \mathbf{Y} is replaced by $\tilde{\mathbf{Y}}$ in (1). Note that $\tilde{\mathbf{Y}}$ is a homogeneous Poisson process on W_{ext} with intensity κ , and we think of W_{ext} as an extended window which should be chosen large enough to ensure that with a high probability the Voronoi tessellation of the observation window W is unchanged if \mathbf{Y} is replaced by $\tilde{\mathbf{Y}}$, meaning that $C_i \cap W = \tilde{C}_i \cap W$ whenever $y_i \in \tilde{\mathbf{Y}}$. Equivalently we want to choose a sufficiently large

enough W_{ext} so that the probability

$$p = \mathbb{P}(C_i \cap W \neq \emptyset \text{ for some } y_i \in \mathbf{Y} \setminus W_{\text{ext}})$$

is small. We have $p \leq M$, where

$$M = \mathbb{E} \sum_i 1[y_i \in \mathbf{Y} \setminus W_{\text{ext}}, C_i \cap W \neq \emptyset]$$

is the expected number of nuclei outside the extended window whose Voronoi cells are intersecting W . Let $\omega_d = \pi^{d/2}/\Gamma(1+d/2)$ be the volume of the unit ball in \mathbb{R}^d , f_d the density of the Gamma-distribution with shape parameter d and scale parameter 1, and

$$c_d = \frac{2^{d+1}\pi^{(d-1)/2}\Gamma((d^2+1)/2)}{d^2\Gamma(d^2/2)} \left(\frac{\Gamma(1+d/2)}{\Gamma(1/2+d/2)} \right)^d.$$

Appendix A verifies the following upper bound given by a one-dimensional integral which can easily be evaluated by numerical methods.

Theorem 1 Let W be included in a closed ball $b(z, r_1)$ with centre z and radius r_1 , and let $W_{\text{ext}} = b(z, r_2)$ where $0 < r_1 \leq r_2 < \infty$. Then $p \leq M \leq N$ where

$$N = \kappa\omega_d c_d \int_{\kappa\omega_d(r_2-r_1)^d}^{\infty} \left(\left[\left(\frac{t}{\kappa\omega_d} \right)^{1/d} + r_1 \right]^d - r_2^d \right) f_d(t) dt.$$

Unfortunately, this theoretical result is too conservative when analyzing the Daisies dataset, where $r_1 = 1.096$ is the smallest possible value: In Fig. 3, the dotted, dashed, and solid lines correspond to N and estimated values of M and p , respectively. The lines are displayed for 20 values of $r_2 \geq r_1$, and the estimated values of M and p are calculated from 100 independent simulations of \mathbf{Y} . When $r_2 \geq 2$, the three curves in Fig. 3 are effectively equal. However, for the Bayesian analysis in Section 4, a much smaller extended window W_{ext} than a ball of radius 2 is in fact sufficient for reducing edge effects to an acceptable level (see Section 4.1 for details on how we specify W_{ext} in the particular case of the Daisies dataset).

3 Models for the offspring density

This section discusses various choices of the offspring density $h(\cdot; C)$ in terms of ‘polar coordinates’ $x = ru$ for $r = \|x\| > 0$ and $u = x/\|x\| \in S_{d-1}$, where S_{d-1} is the unit sphere in \mathbb{R}^d . Hence

$$h(x; C) = f(u; C)f(r|l)/r^{d-1}, \quad x \in \text{int}C \setminus \{o\}, \quad (3)$$

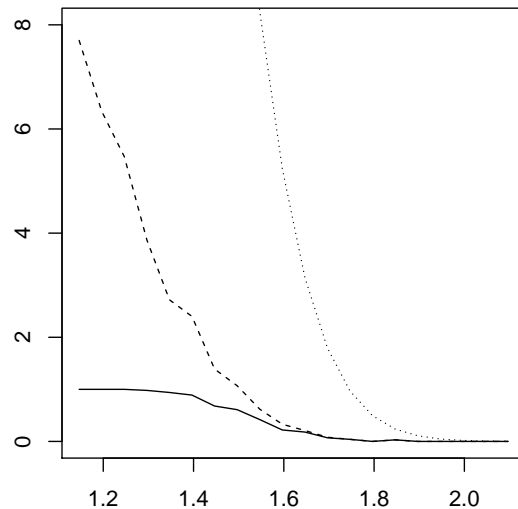


Fig. 3: Results based on Theorem 1 in relation to the Daisies dataset, with z the centre point of W and $r_1 = 1.096$. Upper curve: The upper bound N as a function of $r_2 \geq r_1$. Middle curve: The estimated mean number of nuclei outside the disc $b(z, r_2)$ whose Voronoi cells intersect the disc $b(z, r_1)$ (again as a function of $r_2 \geq r_1$). Lower curve: The estimated probability of having a nucleus outside the disc $b(z, r_2)$ whose Voronoi cell intersects the disc $b(z, r_1)$.

where $f(\cdot; C)$ is a density with respect to surface measure on S_{d-1} , $l = l(u, C)$ is the length of the ray $a(u, C) = \{tu : t > 0, tu \in C\}$ (i.e. the line segment with one endpoint at o and in the direction u the other endpoint at the boundary of C), and $f(\cdot|l)$ is a density with respect to Lebesgue measure on $(0, l)$ (meaning that we exclude the Lebesgue nullset where $x = o$ or x belongs to the boundary of C). In all of our examples, $h(x; C)$ only depends on u through $l = l(u, C)$. Then \mathbf{X} is isotropic, i.e. the distribution of \mathbf{X} is invariant under motions in \mathbb{R}^d .

One proposal is to let

$$h(x; C) \propto \exp(-\|x\|^2/(2\sigma^2)), \quad x \in \text{int}C. \quad (4)$$

This is the restriction of the density of $N_d(0, \sigma^2 I)$ (the radially symmetric d -dimensional normal distribution with zero mean and variance $\sigma^2 > 0$) to $\text{int}C$. The parameter $\sigma > 0$ controls the degree of clustering of the secondary points around the primary points. Let

$$k_0(l) = \int_0^l r^{d-1} \exp(-r^2/(2\sigma^2)) dr$$

which can be expressed in terms of the distribution function for a gamma distribution, e.g.

$$k_0(l) = \sigma^2 [1 - \exp(-l^2/(2\sigma^2))] \quad \text{if } d = 2.$$

Then, by (3) and (4),

$$f(u; C) \propto k_0(l) \quad (5)$$

and

$$f(r|l) = r^{d-1} \exp(-r^2/(2\sigma^2)) / k_0(l), \quad 0 < r < l. \quad (6)$$

The normalizing constant in (5) depends on the boundary of C in a rather complicated way if $d \geq 2$, but at least simulation from $h(\cdot; C)$ can be done by rejection sampling, with $N_d(0, \sigma^2 I)$ as the instrumental distribution.

Another proposal, which avoids the problem above of calculating a normalizing constant, and still with clustering of the secondary points around the primary points, is given by

$$h(x; C) = \frac{l^d \lambda \exp(-\lambda r^d)}{|C| [1 - \exp(-\lambda l^d)]}, \quad x \in \text{int}C. \quad (7)$$

The parameter $\lambda > 0$ controls the degree of clustering; large values of λ corresponds to dense clustering. For $d = 2$, (7) appeared in Møller and Rasmussen (2012) in the context of a sequential point process setting. By (3) and (7),

$$f(u; C) = l^d / [d|C|] \quad (8)$$

and

$$f(r|l) = \frac{dr^{d-1} \lambda \exp(-\lambda r^d)}{1 - \exp(-\lambda l^d)}, \quad 0 < r < l. \quad (9)$$

Note that (9) corresponds to an exponential distribution for $s = r^d$ restricted to the interval $(0, l^d)$, and (9) agrees with (6) if $d = 2$ and $\lambda = 1/(2\sigma^2)$. Simulation from (8)-(9) can simply be done by generating a uniform point z on C and returning $u = z/\|z\|$, and by generating a uniform number $t \in (0, 1)$ and returning

$$s = -(1/\lambda) \log [1 - t (1 - \exp(-\lambda l^d))].$$

To obtain models with clustering of the secondary points around the boundaries of the Voronoi cells, in the right hand expressions of (6) and (9) we may simply replace r by $l - r$. As we need a more flexible model in Section 4, we consider instead a modification where we are still using (8) but replaces (9) by a density such that $s = r/l$ follows a Beta(a, b)-distribution and is independent of l . Then ‘looking along’ the scaled ray $a(u, C_i - y_i)/l(u, C_i - y_i)$ relative to each nucleus y_i and in the direction u , we have clustering of the secondary points

- (a) away from both the nuclei and the Voronoi edges if $a > 1$ and $b > 1$,
- (b) away from the nuclei and around the Voronoi edges if $a \geq 1$, $0 < b \leq 1$, and $(a, b) \neq (1, 1)$,

- (c) around the nuclei and away from the Voronoi edges if $0 < a \leq 1$, $b \geq 1$, and $(a, b) \neq (1, 1)$,
- (d) around the nuclei and the Voronoi edges if $0 < a < 1$ and $0 < b < 1$,

while we have uniformity if $a = b = 1$ (not meaning that $h(\cdot; C)$ is uniform on C —that is the case $a = d$ and $b = 1$). The kind of clustering described in (a) and (d) are not covered by the previous models. As we shall see in Section 4, for the Daisies dataset, (a) becomes the relevant case.

4 Statistical inference

Statistical inference based on maximum likelihood or maximum composite likelihood (see Møller and Waagepetersen (2004, 2007) and the references therein) are complicated by the fact the distribution of the secondary process is intractable: Due to unobserved primary point process, the density of the primary point process is not expressible on closed form; and because of the complicated stochastic structure for the Poisson-Voronoi tessellation, second- and higher moment properties of the secondary point process \mathbf{X} are also infeasible to calculate (see Appendix B). In fact the only simple analytic result we are aware of is the intensity for \mathbf{X} ; if e.g. k is the identity mapping, then \mathbf{X} has intensity $\alpha + \beta$.

In this section we propose instead a Bayesian Markov chain Monte Carlo (MCMC) approach, paying attention to the Daisies dataset in Fig. 1. The missing data, i.e. the primary point process approximated by a Poisson process defined on a sufficiently large region as discussed in Section 2, will be included in the posterior. Thereby we can estimate properties of both the Voronoi cells and the parameter vector $\theta = (\kappa, \alpha, \beta, a, b)$.

4.1 Model specification

Denote the observed point pattern of daisies by $\mathbf{x} = \{x_1, \dots, x_{n_x}\}$ and the unobserved point pattern of nuclei (or cluster centres) by $\mathbf{y} = \{y_1, \dots, y_{n_y}\}$. For the random intensity function Λ in (2), we let k be the identity function, and specify the offspring density $h(x; C)$ in (3) by letting $f(u; C)$ be given by (8), and $f(r|l)$ be the density of a scaled Beta(a, b)-distribution as explained at the end of Section 3. An argument for using this specific model for Λ is that (i) when k is the identity function, each Voronoi cell C_i has a mean number of secondary points (daisies) proportional to its area A_i , and (ii) when using a scaled distribution for $f(r|l)$, the region ‘covered by the daisies’ in C_i is also scaled

proportionally to A_i . This means that small and large clusters of daisies have relatively the same density of points.

We let W be the 1.6×1.5 rectangular window in Fig. 1, and W_{ext} be a rectangular extended window with side lengths that are 1.25 as large as the side lengths of W (W_{ext} has the same center as W). Note that we have also tried using a W_{ext} equal to W or a W_{ext} scaled by a factor 1.5. The results reported in this paper for the scale factor 1.25 are very similar to those obtained when using the scale factor 1.5, suggesting that a factor of 1.25 is adequate for dealing with edge effects.

With our model thus defined, the unobserved data \mathbf{y} follows a Poisson process on W_{ext} with intensity κ , and conditionally on \mathbf{y} the observed data \mathbf{x} follows a Poisson process on W with the intensity function given by (2) (except that the sum in (2) is now over the points in \mathbf{y}). Therefore we have a simple expression for the (prior) density function for \mathbf{y} conditional on θ , namely

$$p(\mathbf{y}|\kappa) = \kappa^{n_y} \exp((1 - \kappa)|W_{\text{ext}}|),$$

where the density is with respect to the unit rate Poisson process on W_{ext} (see e.g. Møller and Waagepetersen (2004)). Further, the density for \mathbf{x} conditional on \mathbf{y} and θ is

$$p(\mathbf{x}|\alpha, \beta, a, b, \mathbf{y}) = \exp\left(|W| - \int_W \Lambda(x) dx\right) \prod_{i=1}^{n_x} \Lambda(x_i),$$

where the density is with respect to the unit rate Poisson process on W . Here

$$\int_W \Lambda(x) dx = \alpha + \beta \sum_{i=1}^{n_y} \int_{W \cap C_i} A_i h(x - y_i; C_i - y_i) dx$$

where we encounter the problem that the latter integral cannot always be analytically solved: This integral is one if $C_i \subset W$, zero if $C_i \cap W = \emptyset$, and otherwise we simulate points from $h(\cdot - y_i; C_i - y_i)$ a number of times and use the relative frequency of points falling in $C_i \cap W$ as a Monte Carlo estimate of the integral.

In addition, in order to obtain the posterior, we also need to equip θ with a (hyper) prior. We use independent uniform priors for the parameters $\kappa, \alpha, \beta, a, b$ restricted to large but bounded intervals, i.e. θ has density

$$p(\theta) \propto \mathbf{1}[\kappa \in I_1, \alpha \in I_2, \beta \in I_3, a \in I_4, b \in I_5],$$

where I_1, \dots, I_5 are intervals starting at zero and ending at large but finite values c_1, \dots, c_5 (see Section 4.2).

Therefore the joint posterior distribution of the parameters and the missing data is given by

$$\begin{aligned} p(\theta, \mathbf{y}|\mathbf{x}) &\propto p(\theta)p(\mathbf{y}|\kappa)p(\mathbf{x}|\alpha, \beta, a, b, \mathbf{y}) \\ &\propto \mathbf{1}[\kappa \in I_1, \alpha \in I_2, \beta \in I_3, a \in I_4, b \in I_5] \\ &\kappa^{n_y} \exp(-\kappa|W_{\text{ext}}|) \exp\left(-\int_W \Lambda(x) dx\right) \prod_{i=1}^{n_x} \Lambda(x_i). \end{aligned} \quad (10)$$

4.2 Markov chain Monte Carlo approach

As the posterior (10) is analytically intractable, we use MCMC methods for inference. Specifically we use a Metropolis-within-Gibbs algorithm (see e.g. Gilks et al. (1996)), where we alternate between updating the parameters and the missing data. For the parameter updates we use Metropolis updates with normally distributed proposals, and for the missing data we use birth/death/move updates (see Geyer and Møller (1994) and Møller and Waagepetersen (2004)).

Note that in the simulation algorithm we do not have to distinguish between cluster and background points because of the simple additive form of (2). However, we could have viewed the two processes separately, and considered the type of points in \mathbf{x} (i.e. the two types of cluster and background points) as additional missing data. Thus, extending our algorithm to this situation, we would have to add updates for the type of points. This would complicate and slow down the simulation procedure. As the type of points is not our primary concern, we have not implemented this extension.

For the posterior analysis of the Daisies dataset, we let the Markov chain run for 100000 steps, discarding the first 10000 steps as burn-in. Trace plots of the five parameters and the number of points have been used to assess that the chain has converged approximately to the target distribution at the burn-in, and that the mixing properties of the algorithm are sufficiently good so that the 90000 steps after the burn-in is appropriate when obtaining MCMC estimates.

It turns out that even with infinite values of the constants c_1, \dots, c_5 , the MCMC algorithm converges, indicating that the posterior is proper even when the prior is an improper uniform distribution. For this reason we avoid choosing specific values for c_1, \dots, c_5 , and merely think of these as being sufficiently large such that they do not influence the MCMC run.

4.3 Data Analysis

Fig. 4 shows the marginal posterior distributions approximated from the MCMC runs. These distributions

are clearly different from the uniform priors, indicating that the posterior results are ‘driven by the data and not the prior’. The corresponding posterior mean estimate of θ is

$$(\hat{\kappa}, \hat{\alpha}, \hat{\beta}, \hat{a}, \hat{b}) = (2.312, 8.233, 17.80, 5.068, 11.18). \quad (11)$$

These values can be interpreted in the following way:

- We expect around $|W|\hat{\kappa} \approx 5.549$ nuclei/centre of clusters in W .
- Roughly two-thirds ($\frac{\hat{\beta}}{\hat{\alpha}+\hat{\beta}} \approx 0.684$) of the points are cluster points, while the rest are background points.
- Since $\hat{a} > 1$ and $\hat{b} > 1$, the cluster points are concentrated away from the nuclei/centre of clusters and from the boundaries of the cells, cf. (a) in Section 3. Furthermore, since $\hat{a} < \hat{b}$, the cluster points tend to be closer to their centres than the Voronoi cell boundaries.

Since the posterior also contains information on the missing data \mathbf{y} , we can estimate where the Voronoi cell boundaries typically are located, giving us an idea of how the clusters are separated. This is illustrated in Fig. 5, which we have obtained in the following way: We extract 100 point patterns from the MCMC runs sampled at regular intervals on the 90000 steps remaining after discarding the burn-in. Denote the collection of these point patterns by $\tilde{\mathbf{y}}$ and the corresponding collection of parameters by $\tilde{\theta}$. We consider the line segment pattern consisting of the union of Voronoi boundaries obtained from every point pattern \mathbf{y}_j in $\tilde{\mathbf{y}}$. The figure shows a kernel estimate of the intensity of line segments obtained by taking the convolution of the line segments and a Gaussian kernel. It is clear from the figure that the clusters of daisies are well-separated by the model. Furthermore, small separations in the clusters are visible as faint light grey. Not surprisingly, the separations seem more random outside W where we have no data.

Another way of visually illustrating the clusters is to summarize posterior results for the intensity $\Lambda(x)$ as follows. For each pair (θ_j, \mathbf{y}_j) in $(\tilde{\theta}, \tilde{\mathbf{y}})$, we calculate the intensity, say $\Lambda_j(x)$, using (2). Fig. 6 shows the estimated posterior mean and variance of the intensity. As expected the mean shows us that the intensity is high where the data has many points. The variance is also higher in regions with many points, and in general shows more artifacts from the samples (i.e. faint light greys resulting from clusters in the model that have only existed for a short while in the Markov chain).

Although we did not include the type of a point (cluster or background) in the posterior, we can still estimate the probability that a point is a cluster point, in a similar manner as we estimated the intensity: From the intensities Λ_j obtained before, we can calculate $1 -$

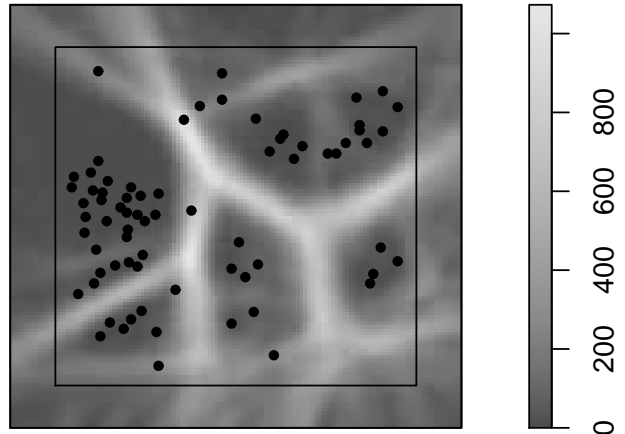


Fig. 5: The posterior distribution of Voronoi boundaries illustrated by their intensity, where the light grey signifies a high intensity (see the text for details). The windows W_{ext} and W are shown as rectangles.

$\alpha_j/\Lambda_j(x)$, where α_j is the α -parameter from θ_j . This is an estimate of the probability that a point located at position x is a cluster point. Fig. 7 shows the mean and variance of this probability for $x \in W_{\text{ext}}$. As expected the mean plot shows us that points that are located in the visual clusters have a high probability of being cluster points, while solitary points have a low probability. The variance has low values in the middle of the visual clusters and far away from the visual clusters, indicating that only points that lie on the border of the visual clusters are hard to classify correctly.

4.4 Model check

Finally, we need to check how well the model actually fits the Daisies dataset. Firstly we do a simple model check by comparing five posterior predictive simulations of the model with the data. These are shown in Fig. 8. Comparing the data and the simulations, we can see no systematic deviations, indicating that the model is producing patterns that are not visually discernible from the data.

Next, for assessing the fit of the model, we consider non-parametric estimates of the summary statis-

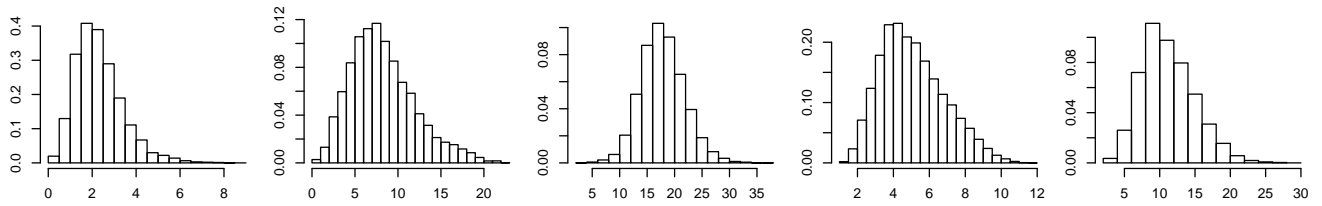


Fig. 4: Marginal posterior distributions of the parameters κ , α , β , a and b approximated by MCMC.

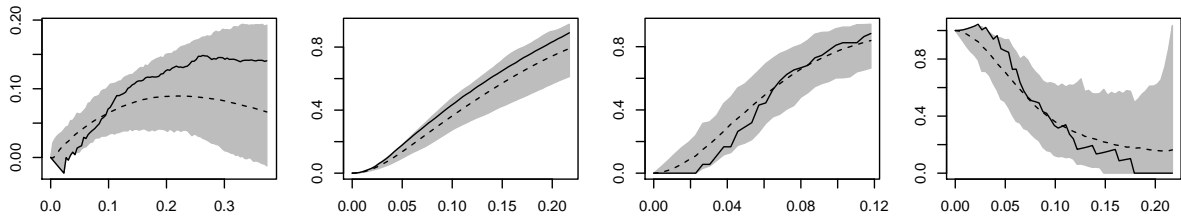


Fig. 9: The L (minus identity), F , G and J -functions estimated from the data (solid lines) shown with 95% pointwise bounds (grey area) and mean (dashed line) calculated from 199 mean simulated datasets when θ is given by its estimated posterior mean.

tics L , F , G and J (for definitions, see e.g. Møller and Waagepetersen (2004)). Here we use the mean posterior estimate (11) of θ when simulating 199 new datasets from which we compute approximate 95% bounds and average of each of the four summary statistics to compare with estimates based on the data. The results are shown in Fig. 9. All of the summary statistics estimated from the data seem to agree well with the summary statistics based on the model. One small point to notice is that Fig. 9 indicates a bit regularity of point pairs at very short ranges (essentially there is a minimum range between neighbouring daisies). This aspect is not included in our model, but the effect of this is very minor.

Finally, for a posterior predictive check of the model, we use the inhomogeneous K -function (Baddeley, Møller and Waagepetersen, 2000) as follows. For each (θ_j, \mathbf{y}_j) in $(\tilde{\theta}, \tilde{\mathbf{y}})$ we calculate non-parametric estimates $\hat{K}_{\theta_j, \mathbf{y}_j, \mathbf{x}}(r)$ of the inhomogeneous K -function based on the Daisies dataset \mathbf{x} , and using the intensity A_j as previously defined. For comparison we simulate new data \mathbf{x}_j for each parameter θ_j and primary process \mathbf{y}_j , and based on these we calculate non-parametric estimates $\hat{K}_{\theta_j, \mathbf{y}_j, \mathbf{x}_j}(r)$ for each j . We then calculate the difference

$$\Delta K_j(r) = \hat{K}_{\theta_j, \mathbf{y}_j, \mathbf{x}}(r) - \hat{K}_{\theta_j, \mathbf{y}_j, \mathbf{x}_j}(r), \quad r > 0,$$

for each j . Fig. 10 shows the mean and 95% bounds of these differences. The zero function is completely inside

the bounds, indicating no discrepancy between the data and the model.

Acknowledgements Supported by the Danish Council for Independent Research—Natural Sciences, grant 12-124675, ”Mathematical and Statistical Analysis of Spatial Data”, and by the Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation.

Appendix A: Proof of Theorem 1

Let the situation be as in Section 2. By the Slivnyak-Mecke Theorem (Møller and Waagepetersen (2004) and the references therein), M is equal to

$$\kappa \int_{W_{\text{ext}}^c} \mathbb{P}(C(y; \mathbf{Y} \cup \{y\}) \cap W \neq \emptyset) dy$$

which by stationarity of \mathbf{Y} can be rewritten as

$$\kappa \int_{W_{\text{ext}}^c} \mathbb{P}(C(o; \mathbf{Y} \cup \{o\}) \cap W_{-y} \neq \emptyset) dy \quad (12)$$

where $W_{\text{ext}}^c = \mathbb{R}^d \setminus W_{\text{ext}}$ is the complement of W_{ext} and W_{-y} is W translated by $-y$. Note that $C(o; \mathbf{Y} \cup \{o\})$ is the so-called typical Poisson-Voronoi cell (see e.g. Møller (1989)). Denoting T the distance from o to the furthest vertex of $C(o; \mathbf{Y} \cup \{o\})$ and $d(o, W_{-y})$ the distance from o to W_{-y} (which is well-defined if e.g.

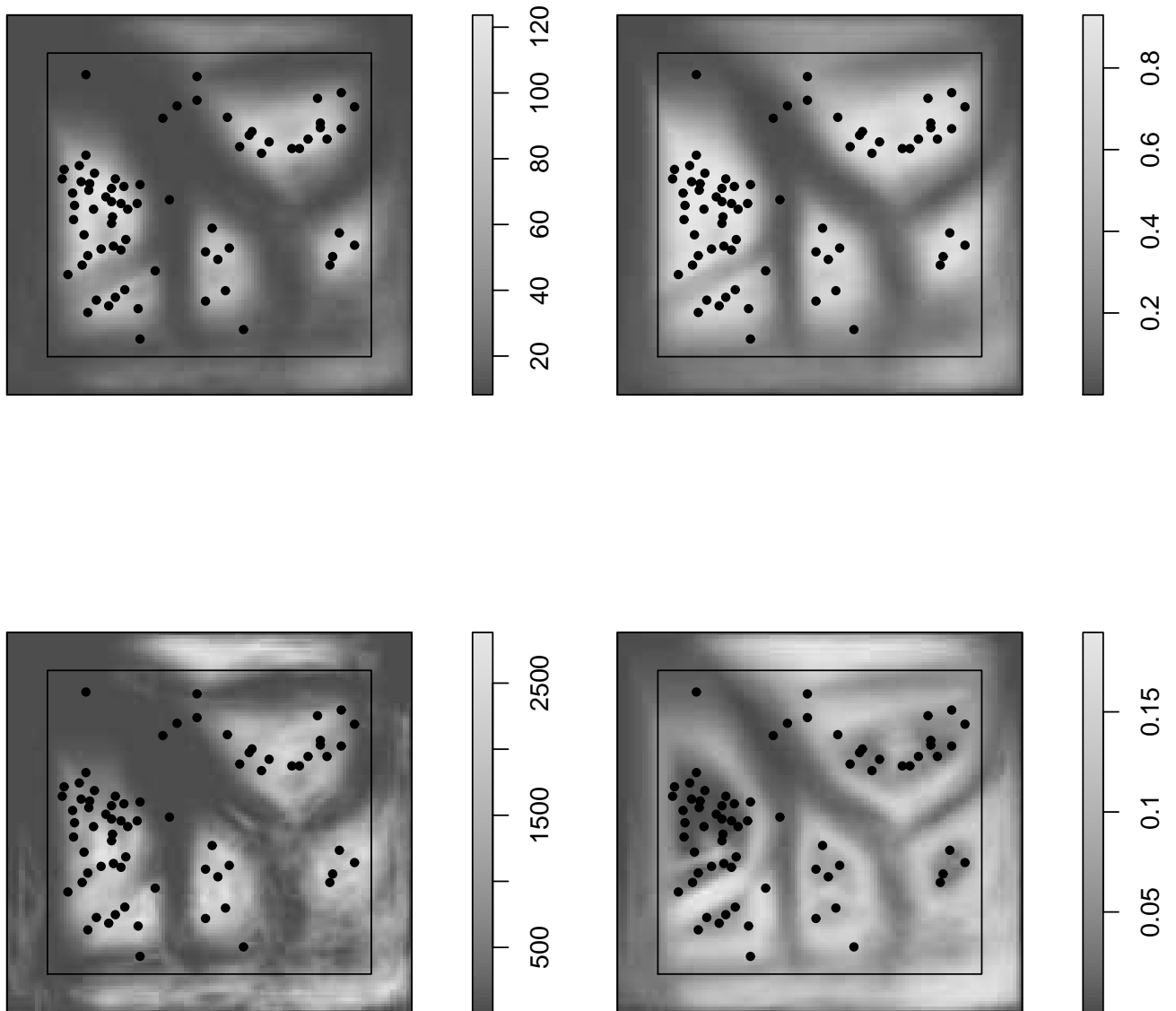


Fig. 6: Upper plot: Posterior mean of $\Lambda(x)$. Lower plot: Posterior variance of $\Lambda(x)$. The windows W_{ext} and W are shown as rectangles in both plots.

Fig. 7: Posterior mean (upper) and variance (lower) of the function $1 - \alpha/\Lambda$. The windows W_{ext} and W are shown as rectangles.

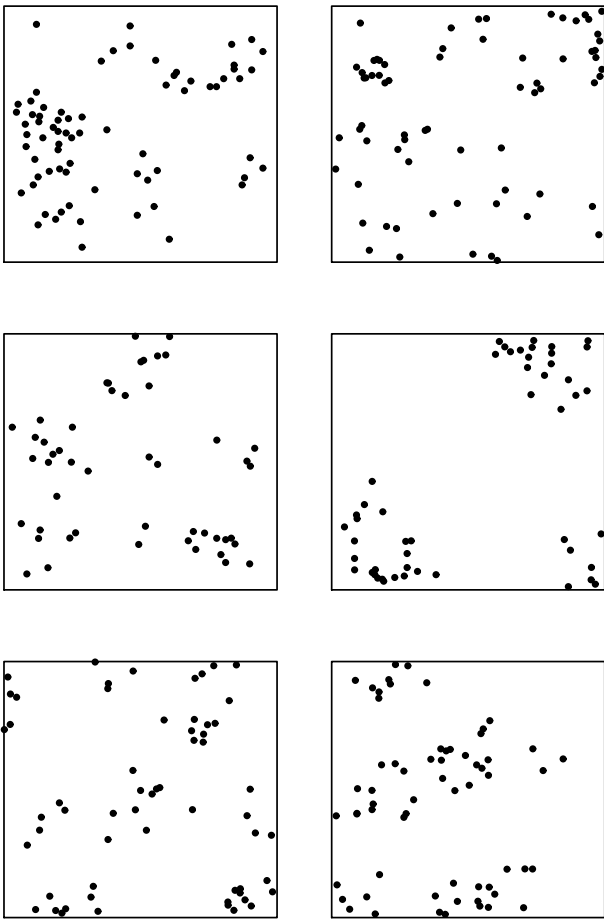


Fig. 8: The upper left point pattern is the Daisies data, while the other five point patterns are posterior predictive simulations.

W_{-y} is compact),

$$\kappa \int_{W_{\text{ext}}^c} \mathbb{P}(T > d(o, W_{-y})) \, dy \quad (13)$$

is an upper bound for (12) and hence also for p .

In order to bound (13), we start by deriving a lower bound on the cumulative distribution function (cdf) of T . Denote $\sigma_d = 2\pi^{d/2}/\Gamma(d/2)$ the surface area of the unit ball in \mathbb{R}^d , and F_d the cdf of the Gamma-distribution with shape parameter d and scale parameter 1.

Lemma 1 If W is compact, then

$$\mathbb{P}(T > t) \leq c_d [1 - F_d(\kappa\omega_d t^d)], \quad t \geq 0. \quad (14)$$

Proof of Lemma 1: We shall ignore nullsets. With prob-

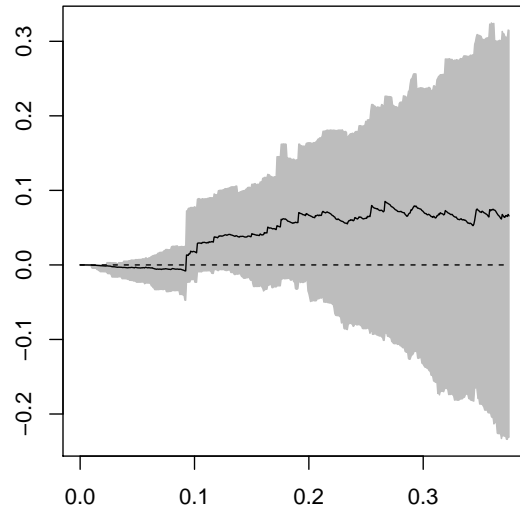


Fig. 10: The difference of inhomogeneous K -functions calculated from data and simulations shown by their mean (solid line) and 95% bounds (grey area). The zero function is indicated by a dashed line.

ability one, for all pairwise distinct points $y_1, \dots, y_d \in \mathbf{Y}$, the d -dimensional closed ball $B(o, y_1, \dots, y_d)$ containing o, y_1, \dots, y_d in its boundary is well-defined. Denote $R(o, y_1, \dots, y_d)$ the radius of $B(o, y_1, \dots, y_d)$. Then $\mathbb{P}(T > t)$ is at most

$$\frac{1}{d!} \mathbb{E} \sum_{\substack{\neq \\ y_1, \dots, y_d \in \mathbf{Y}}} \mathbb{1}[B(o, y_1, \dots, y_d) \cap \mathbf{Y} \setminus \{y_1, \dots, y_d\} = \emptyset, \\ R(o, y_1, \dots, y_d) > t]$$

where \neq over the summation sign means that y_1, \dots, y_d are pairwise distinct, and noting that the sum is almost surely $d!$ times the number of vertices in $C(o; \mathbf{Y} \cup \{o\})$ with distance at least t to o . Therefore, by repeated use of the Slivnyak-Mecke theorem, $\mathbb{P}(T > t)$ is at most

$$\frac{\kappa^d}{d!} \int \dots \int \mathbb{P}(B(o, y_1, \dots, y_d) \cap \mathbf{Y} = \emptyset, \\ R(o, y_1, \dots, y_d) > t) \, dy_1 \dots dy_d$$

and hence, since \mathbf{Y} is a stationary Poisson process and $B(o, y_1, \dots, y_d)$ has volume $\omega_d R(o, y_1, \dots, y_d)^d$, $\mathbb{P}(T > t)$ is at most

$$\frac{\kappa^d}{d!} \int \dots \int \mathbb{1}[R(o, y_1, \dots, y_d) > t] \\ \exp(-\kappa\omega_d R(o, y_1, \dots, y_d)^d) \, dy_1 \dots dy_d \\ = \frac{\kappa^d}{d!} \frac{1}{|A|} \int \int \dots \int \mathbb{1}[y_0 \in A, R(y_0, y_1, \dots, y_d) > t] \\ \exp(-\kappa\omega_d R(y_0, y_1, \dots, y_d)^d) \, dy_0 dy_1 \dots dy_d$$

where $A \subset \mathbb{R}^d$ is an arbitrary Borel with volume $0 < |A| < \infty$, and where $R = R(y_0, y_1, \dots, y_d)$ is the radius of the d -dimensional closed ball $B(y_0, y_1, \dots, y_d)$ containing y_0, y_1, \dots, y_d in its boundary (which is well-defined for Lebesgue almost all $(y_0, y_1, \dots, y_d) \in \mathbb{R}^{d(d+1)}$). Denote $z = z(y_0, y_1, \dots, y_d)$ the centre of $B(y_0, y_1, \dots, y_d)$, $u_i = u_i(y_0, y_1, \dots, y_d)$ the unit vector such that $y_i = z + Ru_i$ ($i = 0, 1, \dots, d$), $\Delta(u_0, u_1, \dots, u_d)$ the volume of the convex hull of u_0, u_1, \dots, u_d , and ν surface measure on the unit sphere in \mathbb{R}^d . Then, by Blasche-Petkantschin's formula (e.g. Miles (1971)), $P(T > t)$ is at most

$$\frac{\kappa^d}{|A|} \int \int \int \int \cdots \int 1[z + Ru_0 \in A, R > t] R^{d^2-1} \exp(-\kappa\omega_d R^d) \Delta(u_0, u_1, \dots, u_d) dz dR \nu(du_0) \nu(du_1) \cdots \nu(du_d)$$

which reduces to

$$\kappa^d \int_t^\infty R^{d^2-1} \exp(-\kappa\omega_d R^d) dR \int \int \cdots \int \Delta(u_0, u_1, \dots, u_d) \nu(du_0) \nu(du_1) \cdots \nu(du_d).$$

Thereby, since

$$\int_t^\infty R^{d^2-1} \exp(-\kappa\omega_d R^d) dR = \frac{(d-1)!}{d(\kappa\omega_d)^d} [1 - F(d(\kappa\omega_d t^d))]$$

and

$$\int \int \cdots \int \Delta(u_0, u_1, \dots, u_d) \nu(du_0) \nu(du_1) \cdots \nu(du_d) = \frac{2^{d+1} \pi^{(d^2+d-1)/2} \Gamma((d^2+1)/2)}{d! \Gamma(d^2/2) \Gamma((d+1)/2)^d}$$

(see Theorem 2 in Miles (1971)), we obtain (14) after a straightforward calculation.

Proof of Theorem 1: It suffices to consider the case where $W = b(z, r_1)$. Then p is at most

$$\kappa \int_{\|z-y\| \geq r_2} P(T > d(o, b(z-y, r_1))) dy \leq \kappa c_d \int_{\|y\| > r_2} \int_{\kappa\omega_d(\|y\|-r_1)^d}^\infty f_d(t) dt dy$$

where the inequality follows from Lemma 1. Hence, using Fubini's theorem, a shift for y to hyperspherical coordinates in \mathbb{R}^d , and the fact that $\omega_d = \sigma_d/d$, we easily deduce the result.

Appendix B: Moment results

Since \mathbf{X} is a Cox process driven by (2), moment results for \mathbf{X} are inherited from the distribution of the primary point process \mathbf{Y} . In particular, \mathbf{X} has intensity $\rho = E\Lambda(o)$ and pair correlation function $g(x) = E[\Lambda(o)\Lambda(x)]/\rho^2$, $x \in \mathbb{R}^d$ (provided these expectations exist), see e.g. Møller and Waagepetersen (2004). This appendix discusses the expressions of ρ and g .

Recall the notion of the typical Voronoi cell: Let Π denote the space of compact convex polytopes $C \subset \mathbb{R}^d$ with $|C| > 0$ and $o \in \text{int}C$ (we equip Π with the usual σ -algebra for closed subsets of \mathbb{R}^d restricted to Π , i.e. the σ -algebra generated by the sets $\{C \in \Pi : C \cap K = \emptyset\}$ for all compact $K \subset \mathbb{R}^d$). The typical Voronoi cell is a random variable \mathcal{C} with state space Π and distribution

$$P(\mathcal{C} \in F) = E \sum_i 1[y_i \in B, C_i - y_i \in F] / (\kappa|B|) \quad (15)$$

where $B \subset \mathbb{R}^d$ is an arbitrary (Borel) set with $0 < |B| < \infty$ (by stationarity of \mathbf{Y} , the right hand side in (15) does not depend on the choice of B). Intuitively, \mathcal{C} is a randomly chosen cell viewed from its nucleus; formally, (15) is the Palm distribution of a Voronoi cell. It follows by standard measure theoretical considerations that

$$E \sum_i f(y_i, C_i - y_i) = \kappa E \int f(y, \mathcal{C}) dy \quad (16)$$

for any nonnegative (measurable) function f , and letting $\mathcal{A} = |\mathcal{C}|$, then $E\mathcal{A} = 1/\kappa$. See e.g. Møller (1989).

Since \mathbf{Y} is a stationary Poisson process, by the Slivnyak-Mecke formula (see e.g. Møller and Waagepetersen (2004)),

$$E \sum_i f(y_i, \mathbf{Y}) = \kappa E \int f(y, \mathbf{Y} \cup \{y\}) dy \quad (17)$$

for any nonnegative (and measurable) function f . By (16)-(17) and stationarity of \mathbf{Y} , we can then take

$$\mathcal{C} = C(o; \mathbf{Y} \cup \{o\}). \quad (18)$$

Proposition 1 If $E k(\mathcal{A}) < \infty$, then

$$\rho = \alpha + \beta \kappa E k(\mathcal{A}) \quad (19)$$

is finite.

Proof: By (2),

$$E\Lambda_{\text{ext}}(o) = \alpha + \beta E \sum_i 1[-y_i \in C_i - y_i] k(A_i) h(-y_i, C_i - y_i).$$

Combining this with (16) and the facts that $\rho = \mathbb{E}A_{\text{ext}}(o)$ and $|C_i| = |C_i - y_i|$, we obtain that ρ is equal to

$$\begin{aligned} &= \alpha + \beta \mathbb{E} \sum_i 1[-y_i \in C_i - y_i] k(|C_i - y_i|) h(-y_i, C_i - y_i) \\ &= \alpha + \beta \kappa \mathbb{E} \int 1[-y \in C] k(\mathcal{A}) h(-y_i, C) dy \\ &= \alpha + \beta \kappa \mathbb{E} k(\mathcal{A}) \end{aligned}$$

whereby the assertion follows.

The pair correlation function g is more complicated to evaluate. For example, let k be the identity function. Then by similar arguments as in the proof of Proposition 1 and by using (17) and (18), we obtain

$$\begin{aligned} &\mathbb{E}[A(o)A(x)] \\ &= \alpha^2 + 2\alpha\rho + \beta^2\kappa \mathbb{E} \int 1[\{y, x + y\} \subset C] |C|^2 h(y, C) \\ &\quad h(x + y, C) dy + (\beta\kappa)^2 \mathbb{E} \int \int \\ &\quad 1[o \in C(y_1; \mathbf{Y} \cup \{y_1, y_2\}), x \in C(y_2; \mathbf{Y} \cup \{y_1, y_2\})] \\ &\quad |C(y_1; \mathbf{Y} \cup \{y_1, y_2\})| |C(y_2; \mathbf{Y} \cup \{y_1, y_2\})| \\ &\quad h(-y_1; C(y_1; \mathbf{Y} \cup \{y_1, y_2\}) - y_1) \\ &\quad h(x - y_2; C(y_2; \mathbf{Y} \cup \{y_1, y_2\}) - y_2) dy_1 dy_2. \end{aligned}$$

Here the first mean value corresponds to the case where two secondary points with locations o and x belong to the same cell, and the second mean value corresponds to the case where they belong to two different cells. We are not aware of any analytic methods for evaluating these mean values, even if $h(\cdot; C)$ is uniform on C , in which case

$$\begin{aligned} &\mathbb{E}[A(o)A(x)] \\ &= \alpha^2 + 2\alpha\rho + \beta^2\kappa \int P(\{y, x + y\} \subset C(o; \Phi \cup \{o\})) dy + \\ &\quad (\beta\kappa)^2 \int \int P(o \in C(y_1; \mathbf{Y} \cup \{y_1, y_2\}), \\ &\quad x \in C(y_2; \mathbf{Y} \cup \{y_1, y_2\})) dy_1 dy_2 \\ &= \alpha^2 + 2\alpha\rho + \beta^2\kappa \int \exp(-\kappa|b(o, \max\{\|y\|, \|x + y\|\})|) \\ &\quad dy + (\beta\kappa)^2 \int \int 1[\|y_1\| \leq \|y_2\|, \|y_2 - x\| \leq \|y_1 - x\|] \\ &\quad \exp(-\kappa|b(o, \|y_1\|) \cup b(x, \|y_2 - x\|)|) dy_1 dy_2 \end{aligned}$$

where the integrals may be evaluated by numerical methods.

References

1. Baddeley B, Møller J, Waagepetersen R (2000) Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* 54: 329–350
2. Baddeley A, Turner R (2005). Spatstat: an R package for analyzing spatial point patterns. *J. Statist. Software* 12: 1–42
3. Barr C, Schoenberg FP (2010) On the Voronoi estimator for the intensity of an inhomogeneous planar Poisson process. *Biometrika* 97: 977–984
4. Blackwell PG (2001) Bayesian inference for a random tessellation process. *Biometrics* 57: 502–507
5. Blackwell PG, Møller J (2003) Bayesian analysis of deformed tessellation models. *Adv. Appl. Probab.* 35: 4–26
6. Byers SD, Raftery AE (2002) Bayesian estimation and segmentation of spatial point processes using Voronoi tilings. In: Lawson AG, Denison DGT (eds) *Spatial cluster modelling*, Chapman & Hall/CRC Press, London
7. Cox DR (1955) Some statistical models related with series of events. *J. Roy. Statist. Soc. B* 17: 129–164
8. Foss SG, Zuyev SA (1996) On a Voronoi Aggregative Process Related to a Bivariate Poisson Process. *Adv. Appl. Probab.* 28: 965–981
9. Geyer CJ, Møller J (1994) Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Statist.* 21: 359–373
10. Gilks WR, Richardson S, and Spiegelhalter DJ (1996) *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London
11. Heikkinen J, Arjas E (1998) Non-parametric Bayesian estimation of a spatial Poisson intensity. *Scand. J. Statist.* 25: 435–450
12. Miles RE (1971) Isotropic random simplices. *Adv. Appl. Probab.* 3: 353–382.
13. Møller J (1989) Random tessellations in \mathbf{R}^d . *Adv. Appl. Probab.* 21: 37–73
14. Møller J, Rasmussen JG (2012) A sequential point process model and Bayesian inference for spatial point patterns with linear structures. *Scand. J. Statist.* 39: 618–634
15. Møller J, Waagepetersen RP (2004) *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC, Boca Raton
16. Møller J, Waagepetersen RP (2007) *Modern spatial point process modelling and inference (with discussion)*. *Scand. J. Statist.* 34: 643–711
17. Okabe A, Boots B, Sugihara K, Chiu SN (2000) *Spatial tessellations. Concepts and applications of Voronoi diagrams*, 2nd. Wiley, Chichester
18. Skare Ø, Møller J, Jensen EBV (2007) Bayesian analysis of spatial point processes in the neighbourhood of Voronoi networks. *Statist. Comp.* 17: 369–379