Hierarchical spatial point process analysis for a plant community with high biodiversity

Janine B. Illian[1], Jesper Møller[2], Rasmus P. Waagepetersen[2]

[1]SIMBIOS
University of Abertay
Bell Street, Dundee DD1 1HG
UK
j.illian@abdn.ac.uk

[2]Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7G, DK-9220 Aalborg
Denmark
jm,rw@math.aau.dk

Keywords: Bayesian inference; Ecological plant communities; Maximum likelihood; Multivariate spatial point process; Spatial interaction.

Abstract

A complex multivariate spatial point pattern for a plant community with high biodiversity is modelled using a hierarchical multivariate point process model. In the model, interactions between plants with different post-fire regeneration strategies are of key interest. We consider initially a maximum likelihood approach to inference where problems arise due to unknown interaction radii for the plants. We next demonstrate that a Bayesian approach provides a flexible framework for incorporating prior information concerning the interaction radii. From an ecological perspective, we are able both to confirm existing knowledge on species' interactions and to generate new biological questions and hypotheses on species' interactions.

1 Introduction

1.1 A plant community with high biodiversity

This paper discusses a multivariate spatial point process model for spatial point patterns formed by a natural plant community with a high degree of biodiversity. The data were collected in Cataby, in the Mediterranean type shrub- and heathland of the south western area of Western Australia (Beard, 1984), where the locations of 6378 plants from 67 species on a 22 m by 22 m plot have been recorded (Armstrong, 1991). To overcome the problem of identifying individuals and their locations for strongly clonal plants, methods as described in Magurran (1988) were applied.

Figure 1 shows the point patterns of the 24 most abundant and influential species. The patterns are of varying nature, showing different patterns of inhomogeneity and clustering.

[Figure 1 about here.]

Our objective is to devise methodology suitable for modelling a multivariate spatial point pattern dataset as complex as this. We also aim at emphasising the suitability of spatial point process methods in the context of plant community ecology and the applicability of the methodology to other, similar datasets. In particular, we demonstrate that a Bayesian approach provides a flexible framework for incorporating prior information concerning plant interactions.

1.2 Ecological perspective

The dataset considered here is only one example of a multivariate spatial point pattern of high dimensionality. Similar large datasets are becoming increasingly available in the

context of plant community ecology (Burslem *et al.*, 2001). The ecological information buried in these datasets has not yet been fully explored.

In view of the increasing impact of humans on plant ecosystems, often resulting in the destruction of these and thus in a loss of biodiversity, ecologists seek to understand the processes that organise ecological plant communities. Such knowledge may help the re-naturation and subsequent conservation of disturbed ecosystems (Greig-Smith, 1983; Herben *et al.*, 2000). This aspect is further discussed in Section 2 for our specific dataset.

In particular, revealing the mechanisms that allow a large number of species to coexist is of key interest within community ecology (Murrell *et al.*, 2001; Loreau *et al.*, 2001). Species coexistence is directly linked to local inter- and intra-specific interactions in a community (Durrett and Levin, 1998). Each plant has a dependence on local growing conditions (Stoll and Weiner, 2000) and interacts mainly with its immediate neighbours, as plants are non-motile organisms. As noted, the dataset considered here consists of as many as 67 species, which manage to coexist in the same ancient community without out-competing each other. We are interested in revealing the mechanisms that facilitate this coexistence, i.e. in understanding the nature of the interaction structure in the community in terms of the strength and direction of interaction (Law *et al.*, 1997).

1.3 Statistical methodology

Non-statistical spatially explicit individual based models have become a very popular approach in ecological modelling over the last ten years (Law *et al.*, 2003). These mechanistic models are built on equations capturing those properties of the individuals which are believed to influence community dynamics. However, the individual based models yield simulated realisations which do not necessarily capture overall spatial characteristics of the entire community. Moreover, extending the mechanistic model to a statistical model

which takes uncertainty into account would be a difficult task due to the high complexity of the model and the limited data available when a plant community is only observed at one instance in time.

We propose instead to analyse information on locations of individual plants using a spatial point process model of the interaction structure in the plant community. So far, most applications of spatial point processes in ecology have been either of a merely descriptive nature using second order summary statistics such as Ripley's K-function (Dale *et al.*, 2002; Liebhold and Gurevitch, 2002; Wiegand and Moloney, 2004) or the models were restricted to a very small number of species, typically not more than two or three (Mateu *et al.*, 1998). A traditional spatial point process analysis involving the inspection of first and second order summary statistics for each of the species as well as pairwise cross species summary statistics (see e.g. Diggle, 2003; Møller and Waagepetersen, 2003) becomes a very tedious task with high numbers of species.

For our data set (see Section 2) we are interested in studying the interactions between 5 seeders and 19 resprouters - hence one would need to interpret 95 plots of pairwise cross statistics. It is very hard to get an overall understanding of the community dynamics from these plots since all of them would then have to be considered simultaneously. Moreover, using pairwise cross statistics for two different species it is not possible to adjust for the effect of the remaining species. Instead, we develop a suitable hierarchical spatial point process model where seeder locations are modelled conditional on the resprouter locations. The parameters of this model provide a compact and more easily interpretable representation of the interactions between reseeders and resprouters (Figure 2 and Sections 4.1 and 5.2.1). Further, our approach may be widely applied in contexts other than the current data set of seeders and resprouters. For instance, a similar modelling approach could be envisaged for spatial patterns of annual and perennial plants in moderate climates

with the perennials taking on the role of the resprouters and the annuals that of the seeders. Another possible application is modelling of seedlings in the presence of adult trees (Gratzer *et al.*, 2006; Grabarnik and Särkkä, 2006).

1.4 Outline

Section 2 gives further information on the dataset, providing a justification for the hierarchical model considered in Section 3. The hierarchical model exploits subject matter knowledge concerning post-fire regeneration strategies of the species. Some model parameters reflect strength and direction of interaction between individuals while other model parameters define circular regions around the individuals where interaction takes place. Section 4 discusses maximum likelihood inference and points out various problems with the estimation of interaction radii in the model. Section 5 considers a Bayesian approach which incorporates prior information on the interaction radii and overcomes the estimation problem. Section 6 contains concluding remarks.

2 Biological background to dataset

The flora in the sampled area is distinguished by its large number of species specific to that area, the large number of rare flora and a high biodiversity; indeed the southwest of Australia is one of the world's biodiversity hot-spots (Crawley, 1997). Lying within a mineral sand mining area, the study area was mined shortly after data collection in 1990 and will have to be re-naturated as soon as mining has ceased. Current efforts towards re-naturation in neighbouring areas, however, have resulted in a very small survival rate for some species. Nevertheless, re-naturation is a legal requirement in the area after large scale mining and the modelling approach here seeks to inform that re-naturation process

by identifying community dynamics.

The characteristic sandy soil in the area is extremely low in nutrients and water (Armstrong, 1991). Hence, individual plants may interact in various ways by inhibiting each other's growth while competing for the scarce resources (Richardson *et al.*, 1995). This is likely to reduce the number of plants in close proximity thus leading to an observed spatial pattern with inhibition between plants. Attraction may, however, also occur if conditions (shade, nutrients, water availability) in the micro habitat are made more suitable for a plant by the presence of another plant. Except for small-scale local variations caused by individual plants, the extremely low nutrient and water level in the soil can be regarded as spatially homogeneous. The apparent inhomogeneity in Figure 1 e.g. for seeder 2 and resprouters 8 and 15 is therefore likely to be due to a "patchy" growth behaviour typical for the area rather than to inhomogeneity in the environment (Dixon, 2005).

The vegetation in the area is Banksia low woodland (Elkington, 1991), which is susceptible to regular natural fires occurring approximately every ten years. Plants have adapted to this through the development of two different regeneration strategies. *Resprouting species* survive the fire with the above soil part of the plant burning off and the plant regrowing from sub-soil root stock. *Seeding species*, on the other hand, die in the fire but the fire triggers the shedding of seeds, which have been stored since the previous fire. The specific regeneration method used by a species will have an impact on the pattern formed by the individual plants (Armstrong, 1991) and also on the structure of the interaction between species with different regeneration strategies.

Since the resprouting plants have been at exactly the same location for a very long time (some of them for hundreds or even thousands of years; Armstrong, 1991), the seeders, which start anew after each fire, do so with the resprouters already present. We thus assume that the growth and survival of reseeding plants will be influenced by the

resprouting plants already present in the plot, whereas an influence of the seeders on the resprouting plants is highly unlikely. To capture the asymmetric interaction, we model the locations of the reseeding plants conditionally on the locations of the resprouting plants in Section 3.

In the following, we consider the 24 point patterns in Figure 1, i.e. the 5 most abundant species of seeders and the 19 species of resprouters suspected to be most influential either as a result of their abundance (resprouters 1-13) or their size (resprouters 14-19) (Armstrong, 2005). We assume that each individual resprouter has a circular zone of influence with the strength of influence decreasing with the distance from each plant. Beyond this zone the influence is considered zero. The interaction can be positive (e.g. caused by subcanopy soil enrichment; Callaway, 1995) or negative (e.g. caused by allelopathy; Armstrong, 1991). Prior information (Armstrong, 2005) available on the typical radii of the influence zones is given in Table 1 which also shows the numbering we use for resprouter species. The 5 species of seeders are 1. *Andersonia heterophylla*, 2. *Astroloma xerophyllum*, 3. *Conospermum crassinervium*, 4. *Leucopogon conostephioides* and 5. *Leucopogon striatus*.

[Table 1 about here.]

3 Likelihood

The observation model (i.e. the likelihood) presented below and the thinking behind it resembles approaches derived from ecological field theory introduced in Wu *et al.* (1985) that model competition using influence potentials based on distances between individuals (Kuuluvainen and Pukkala, 1889; Økland *et al.*, 1999; Miina and Pukkala, 2002; Kühlmann-Berenzon *et al.*, 2005).

Denote W the 22 m by 22 m plot where the plants have been recorded, $\mathbf{x}_1, \ldots, \mathbf{x}_{19}$ the observed point patterns for the 19 resprouters, $\mathbf{y}_1, \ldots, \mathbf{y}_5$ the observed point patterns for the 5 seeders, and $\mathbf{X}_1, \ldots, \mathbf{X}_{19}, \mathbf{Y}_1, \ldots, \mathbf{Y}_5$ the corresponding spatial point processes, i.e. here each $\mathbf{X}_j$ or $\mathbf{Y}_i$ is considered to be a random finite subset of W; for details on spatial point processes, see Møller and Waagepetersen (2003) and the references therein. Since we are mainly interested in the inter-species interactions between seeders and resprouters, we leave the marginal distribution of the resprouters unspecified and restrict attention to the conditional likelihood of the seeders given the resprouters.

Conditional on $\mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_{19} = \mathbf{x}_{19}$, we assume that $\mathbf{Y}_1, \ldots, \mathbf{Y}_5$ are independent Poisson processes. For convenience we assume a log linear form of the intensity function of $\mathbf{Y}_i$,

$$\lambda(\xi|\mathbf{x}, \boldsymbol{\theta}_i) = \exp\left(\boldsymbol{\theta}_i s(\xi|\mathbf{x})^{\mathsf{T}}\right), \qquad \xi \in W, \tag{1}$$

where we use the following notation: $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_{19})$ is the collection of all 19 resprouter point patterns; $\boldsymbol{\theta}_i = (\theta_{i0}, \ldots, \theta_{i19})$ is a vector of parameters, where $\theta_{i0} \in \mathbb{R}$ is an intercept and for $j = 1, \ldots, 19$, we let $\theta_{ij} \in \mathbb{R}$ control the influence of the jth resprouter on the ith seeder (a positive value of θ_{ij} means a positive/attractive association; a negative value of θ_{ij} means a negative/repulsive association); $s(\xi|\mathbf{x}) = (1, s_1(\xi|\mathbf{x}), \ldots, s_{19}(\xi|\mathbf{x}))$ with

$$s_j(\xi|\mathbf{x}) = \sum_{\eta \in \mathbf{x}_j} h_\eta(\|\xi - \eta\|), \qquad j = 1, \ldots, 19,$$

where $\|\cdot\|$ denotes Euclidean distance, and we have chosen a simple smooth interaction function

$$h_\eta(r) = \begin{cases} \left(1 - (r/R_\eta)^2\right)^2 & \text{if } 0 < r \leq R_\eta \\ 0 & \text{else} \end{cases}$$

for $r \geq 0$, where $R_\eta \geq 0$ defines the radii of interaction of a given resprouter at location

η, cf. Table 1.

Thus, given the resprouters $\mathbf{x}$ the number N_i of points in Y_i is Poisson distributed with mean value $\int_W \lambda(\xi|\mathbf{x},\boldsymbol{\theta}_i)\mathrm{d}\xi$, and if we also condition on N_i, the points in $\mathbf{Y}_i$ are independent and identically distributed with a density proportional to $\lambda(\xi|\mathbf{x},\boldsymbol{\theta}_i)$. It follows that the conditional log likelihood function based on the 5 seeder point patterns $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_5)$ is

$$l(\boldsymbol{\theta}, \mathbf{R}; \mathbf{y}|\mathbf{x}) = \sum_{i=1}^{5} \Big[\boldsymbol{\theta}_i \sum_{\xi \in \mathbf{y}_i} s(\xi|\mathbf{x})^{\mathsf{T}} - \int_W \exp\left(\boldsymbol{\theta}_i s(\xi|\mathbf{x})^{\mathsf{T}}\right) \mathrm{d}\xi \Big], \tag{2}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_5)$ is the vector of all 100 parameters θ_{ij} and $\mathbf{R}$ is the vector of all 3168 radii R_η, $\eta \in \mathbf{x}_j$, $j = 1, \ldots, 19$. In comparison, there are $N_1 + \cdots + N_5 = 1954$ seeders.

A potential weakness of our model is the assumed conditional independence between the seeders, see Sections 4.2 and 5.2.2 on model assessment.

4 Maximum likelihood estimation

4.1 Assumptions and results

For a fixed value of $\mathbf{R}$, due to the log linear intensity (1), the log likelihood (2) can be maximised relatively easily with respect to $\boldsymbol{\theta}$ using the Berman and Turner (1992) device: briefly, the log likelihood is then approximated by a weighted log likelihood of independent Poisson variables which can be optimised using standard software, such as the function `glm` in `R`, see also the `R` package `spatstat` (Baddeley and Turner, 2005). Maximisation with respect to $\mathbf{R}$ on the other hand is difficult due to the high dimensionality of $\mathbf{R}$ and since the R_η-values do not enter the likelihood function in a log linear fashion. In this

section, we therefore make the simplifying assumption that for each resprouter species j, all R_η with $\eta \in \mathbf{x}_j$ are equal to a common interaction radius R_j, given by the midpoint of the interval for the zone of influence in Table 1.

Given the chosen value of $\mathbf{R}$, another problem is the existence of the maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$. By exponential family results (Barndorff-Nielsen, 1978), the MLE for $\boldsymbol{\theta}_i$ exists if and only if $t_j(\mathbf{y}_i) \equiv \sum_{\xi \in \mathbf{y}_i} s_j(\xi|\mathbf{x}) > 0,\ j = 1, \ldots, 19$. Depending on the value of R_j it may well happen that $t_j(\mathbf{y}_i) = 0$, and indeed this is sometimes the case with our choice of the R_j. In Figure 2, the fields marked with NA correspond to the 18 parameters where the MLE does not exist. We could choose to let $\theta_{ij} = -\infty$ if $t_j(\mathbf{y}_i) = 0$. This corresponds to a hard core effect of the jth resprouter on the ith seeder, i.e. the existence of a zone around each resprouter of type j in which no seeder of type i can survive, cf. resprouter 2, 9, and 14. It may on the other hand also simply be the case that the chosen interaction radii for these resprouters are too small.

To study the strengths and directions of interaction, we consider z-statistics for the various θ_{ij}, i.e. MLEs scaled by the associated estimated standard errors. The upper plot in Figure 2 shows transformed z-statistics $u = \Phi(z)$, where Φ is the standard normal distribution function. Dark shading corresponds to large u-values. Assuming that the z-statistics are approximately standard normal (and hence that the us are approximately uniform on $[0, 1]$), the starred fields correspond to parameters which are significantly different from zero at the 5 % level. With 95 interaction parameters, there is obviously an issue of multiple testing and hence we should be careful when interpreting the "starred" parameters. Rather than looking at individual parameters it seems more appropriate to look for resprouters or seeders with a consistent pattern of dark or light fields in Figure 2. For example, it seems reasonable to conclude that resprouters 1 and 4 have a repulsive effect on the seeders whereas there may be evidence that resprouters 15 and 18 have an

attractive effect on the seeders.

[Figure 2 about here.]

4.2 Model assessment

We assess the fitted model using the so-called L-function derived from the inhomogeneous K-function by $L(r) = \sqrt{K(r)/\pi}$, (Baddeley *et al.*, 2000). An estimate $\hat{L}(r; \mathbf{y}_i, \hat{\boldsymbol{\theta}}_i)$ for the ith seeder is obtained by replacing $K(r)$ with the estimate

$$\hat{K}(r; \mathbf{y}_i, \hat{\boldsymbol{\theta}}_i) = \sum_{\xi,\eta \in \mathbf{y}_i} \frac{\mathbf{1}[0 < \|\xi - \eta\| \leq r]}{\lambda(\xi|\mathbf{x}, \hat{\boldsymbol{\theta}}_i)\lambda(\eta|\mathbf{x}, \hat{\boldsymbol{\theta}}_i)} e_{\xi,\eta},$$

where $\mathbf{1}[\cdot]$ is the indicator function and $e_{\xi,\eta}$ is an edge correction factor (Møller and Waagepetersen, 2003). For a Poisson process, $L(r) - r = 0$. The estimated $(L(r) - r)$-functions with 95% envelopes obtained by simulation under the fitted model (Møller and Waagepetersen, 2003) are shown in Figure 3. For seeder 1, the estimated L-function is clearly above the envelopes for distances up to about 40 cm, indicating that there is clustering present in the point pattern that is not explained by the model. This clustering might be a result of a reproduction mechanism where offspring are located in the vicinity of the parent plant. For seeders 2 and 5, the estimated L-functions are below the envelopes at larger distances which indicates the need for including a spatial trend term, see also Section 2. The model yields a reasonably good fit for seeders 3 and 4, though at very small distances there appears to be some intra-specific repulsion. This might be the effect of a zone around each individual in which no conspecific individuals can survive. Note that when estimated L-functions indicate a lack of fit, it is not possible to conclude decisively whether this is due to a misspecified model for the intensity or due to failure of the Poisson process independence assumptions. In a more elaborate model assessment in Section 5.2.2

we supplement plots based on L-functions with residual plots for quadrat counts.

[Figure 3 about here.]

The results indicate that a more appropriate model would have to take intra-specific interaction into account. Assuming known interaction radii, Högmander and Särkkä (1999) consider a hierarchical model with interactions for a bivariate point pattern of ants' nests. However, in our situation, we believe that the assumption of known and equal interaction radii for resprouters of the same type is highly unrealistic, since the plants vary in size, and we choose to focus on this problem in the next section.

5 Bayesian inference

We now consider a more flexible Bayesian approach, allowing different interaction radii for each resprouter plant, and where we elicit prior distributions based on the information in Table 1.

5.1 Prior and posterior

The prior information concerning ranges of zones of influence in Table 1 is the result of extensive and detailed discussions with the scientist who collected the data. We use this prior information to elicit informative priors on the interaction radii. A priori the interaction radii R_η are assumed to be independent random variables and for each $\eta \in \mathbf{x}_j$, R_η follows the restriction of $\mathrm{N}(\mu_j, \sigma_j^2)$ to $[0, \infty)$, where (μ_j, σ_j^2) is chosen so that under the unrestricted $\mathrm{N}(\mu_j, \sigma_j^2)$, the range of the zone of influence in Table 1 is a central 95% interval. Moreover, given the R_η, the θ_{ij} are assumed to be independent and identically distributed, following a relatively non-informative $\mathrm{N}(0, \sigma^2)$-distribution, where the specification of σ is discussed below.

Our a priori independence assumptions above are essentially made, since we have no prior knowledge on how to specify a correlation structure for all the R_η and all the θ_{ij}. For the same reason, $(\boldsymbol{\theta}, \mathbf{R})$ and $\mathbf{X}$ are assumed to be independent. Hence the posterior density for $(\theta, \mathbf{R})$ satisfies

$$\pi(\boldsymbol{\theta}, \mathbf{R}|\mathbf{x}, \mathbf{y}) \propto \pi(\boldsymbol{\theta}, \mathbf{R}) \exp\left(l(\boldsymbol{\theta}, \mathbf{R}; \mathbf{y}|\mathbf{x})\right).$$

Combining the prior assumptions with the log likelihood (2) we obtain the posterior density

$$\begin{aligned}\pi(\boldsymbol{\theta}, \mathbf{R}|\mathbf{x}, \mathbf{y}) \propto \exp\Big(-\sum_{i=1}^{5} \Big[\theta_{i0}^2/(2\sigma^2) - \sum_{j=1}^{19} \Big\{\theta_{ij}^2/(2\sigma^2) + \sum_{\eta \in \mathbf{x}_j} (R_\eta - \mu_j)^2/(2\sigma_j^2)\Big\}\Big]\Big) \\ \times \exp\Big(\sum_{i=1}^{5} \Big[\boldsymbol{\theta}_i \sum_{\xi \in \mathbf{y}_i} s(\xi|\mathbf{x})^\mathsf{T} - \int_W \exp(\boldsymbol{\theta}_i s(\xi|\mathbf{x})^\mathsf{T}) \mathrm{d}\xi\Big]\Big), \quad \theta_{ij} \in \mathbb{R}, R_\eta \geq 0. \quad (3)\end{aligned}$$

The specification of the prior standard deviation σ for the θ_{ij}'s is difficult. In regression models with a design matrix of full rank, a common choice are flat improper priors, i.e. $\sigma^2 = \infty$. In our situation, however, with improper priors on the $\boldsymbol{\theta}_i$'s we cannot guarantee a proper posterior, since the statistics $t_j(\mathbf{y}_i)$ have positive posterior probability of being zero (Waagepetersen, 2005). It is also difficult to specify informative priors on the $\boldsymbol{\theta}_i$'s, since we only have a qualitative understanding of these parameters. It turns out that essentially the same posterior results are obtained with different values $2, 4, 8, \ldots$ of σ. In the following we restrict attention to the results for $\sigma = 8$.

Monte Carlo estimates of posterior distributions are calculated using simulations from (3). We use a hybrid Markov chain Monte Carlo algorithm (see e.g. Robert and Casella, 1999), where $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_5$ are updated in turn, using random walk Metropolis updates, followed by a random walk Metropolis update of $\mathbf{R}$. The proposal distributions for these

random walk updates are multivariate normal with diagonal covariance matrices. The vector of proposal standard deviations for $\boldsymbol{\theta}_i$ is given by $k\hat{\sigma}_{i|\mathbf{y}}$, where k is a user specified parameter and $\hat{\sigma}_{i|\mathbf{y}}$ is an estimate of the vector of posterior standard deviations for $\boldsymbol{\theta}_i$ obtained from a pilot run. The value of k was chosen to give acceptance rates around 25 %. The vector of proposal standard deviations for $\mathbf{R}$ is given by the vector of prior standard deviations divided by 2.

5.2 Results

5.2.1 Interaction parameters

The lower plot in Figure 2 is a grey scale plot of the posterior probabilities $P(\theta_{ij} > 0|y)$ where dark grey scales are associated with large values of these posterior probabilities. The starred parameters are those for which 0 is outside the 95% posterior interval of θ_{ij} given by the 2.5% and 97.5 % quantiles. In comparison with the upper plot in Figure 2, it is striking that the Bayesian approach seems to yield more clear-cut results than the maximum likelihood inference, since the intermediate grey scales are less frequent in the lower plot. Moreover, more θ_{ij}'s have strong evidence for being different from zero if 'strong evidence' is interpreted as being significant at the 95% level or being outside the 95% posterior interval, respectively. Similar to the maximum likelihood results, resprouter 1 seems to have a clear repulsive effect on seeders, and this also seems to be the case for resprouters 2, 4, 5, 8, 9, and 14 (again we should exercise caution with resprouters 2, 9, and 14 where the prior may be concentrated on too small interaction radii). Resprouters 15 and 18 seem to have a distinct attractive effect on seeders. Looking at rows in the lower plot in Figure 2, seeders 1 and 4, for example, seem to be repulsed by resprouters 1-5 and 8-11 and attracted by resprouters 12, 15, and 18. The Bayesian analysis shows

that it may not be valid to interpret all the θ_{ij}'s with NA's in Figure 2 as corresponding to hard cores, since a number of these θ_{ij}'s do not have strong posterior evidence of being different from zero.

The individual interaction radii are not of particular interest, so for resprouters $j = 1, \ldots, 19$ we have just considered the posterior distributions of the empirical mean $\bar{R}_j$ and the empirical standard deviation s_j for the R_η with $\eta \in \mathbf{x}_j$ (not shown). Except for the very sparse resprouters 16 and 19, there is very little difference between the prior mean or standard deviation and the posterior mean of the empirical mean or standard deviation. Hence the prior assumptions are not contradicted by the data.

5.2.2 Model assessment

In this section we follow the idea of posterior predictive model assessment (Gelman *et al.*, 1996) and compare various summary statistics with their posterior predictive distributions. The posterior predictive distribution of statistics depending possibly both on the points $\mathbf{Y}$ and the parameters $\boldsymbol{\theta}$ and $\mathbf{R}$ is obtained from simulations: we generate a posterior sample $(\boldsymbol{\theta}_{i,1}, \mathbf{R}_1), \ldots, (\boldsymbol{\theta}_{i,m}, \mathbf{R}_m)$, and for each $(\boldsymbol{\theta}_{i,k}, \mathbf{R}_k)$ new data $\mathbf{y}_{i,k}$ from the conditional distribution of $\mathbf{Y}_i$ given $(\boldsymbol{\theta}_{i,k}, \mathbf{R}_k)$. We use $m = 100$ approximately independent simulations obtained by subsampling a Markov chain of length $200,000$.

The grey scale plots in Figure 4 are 'residual' plots based on quadrat counts. We divide the observation window into 100 equally sized quadrats and for each seeder count the number of plants within each quadrat. The grey scales reflect the probabilities that counts drawn from the posterior predictive distribution are less or equal to the observed quadrat counts where dark means high probability. The posterior predictive distribution of the quadrat counts is obtained from posterior predictive samples $\mathbf{y}_{i,k}$, $k = 1, \ldots, 100$, as mentioned above. The stars mark quadrats where the observed counts are 'extreme'

in the sense of being either below the 2.5% quantile or above the 97.5% quantile of the posterior predictive distribution. The plot for seeder 1 indicate a lack of fit due to many 'extreme' counts. However, a 'systematic' discrepancy from the assumed model for the intensity is not obvious and the lack of fit could be caused by clustering due to seed dispersal around parent plants. The residual plots for the other seeders do not provide obvious evidence against our model except perhaps for a small group of adjacent 'extreme' counts for seeder 5.

[Figure 4 about here.]

As in Section 4.2 we also consider plots based on L-functions. Denote by $L(r; \mathbf{Y}_i, \boldsymbol{\theta}, \mathbf{R})$ the estimate of the L function obtained from the point process $\mathbf{Y}_i$ using the intensity function corresponding to the interaction parameter vector $\boldsymbol{\theta}$ and interaction radii $\mathbf{R}$. We then consider the posterior predictive distribution of the differences $\Delta_i(r) = L(r; \mathbf{y}, \boldsymbol{\theta}_i, \mathbf{R}) - L(r; \mathbf{Y}_i, \boldsymbol{\theta}_i, \mathbf{R})$, $r > 0$, i.e. the distribution obtained when $(\mathbf{Y}_i, \boldsymbol{\theta}_i, \mathbf{R})$ are generated under the posterior predictive distribution given the data $\mathbf{y}$. If zero is an extreme value in the posterior predictive distribution of $\Delta_i(r)$ for a range of distances r, we may question the fit of our model. As for the quadrat counts, the posterior predictive distribution is computed from a posterior predictive sample $L(r; \mathbf{y}_i, \boldsymbol{\theta}_{i,k}, \mathbf{R}_k) - L(r; \mathbf{y}_{i,k}, \boldsymbol{\theta}_{i,k}, \mathbf{R}_k)$, $k = 1, \ldots, 100$.

Figure 5 presents estimated upper and lower boundaries of the 95 % posterior intervals for the posterior predictive distributions of $\Delta_i(r)$, $r > 0$, for the 5 seeder species ($i = 1, \ldots, 5$). These intervals take into account the uncertainty of the model parameters $\boldsymbol{\theta}$ and $\mathbf{R}$ and are quite wide in comparison with the envelopes in Figure 3. The wide envelopes probably arise because of the posterior uncertainty regarding the interaction radii; the intensity function at a seeder location may a posteriori be very variable if it is highly uncertain whether the seeder location falls within a resprouter influence zone or not. In

accordance with the results in Section 4.2 and the posterior predictive residual plots, there is evidence of clustering for seeder 1. In contrast with Section 4.2, the posterior predictive intervals for seeder 2 indicate clustering. As for seeder 1, this may be explained by offspring clustering around locations of parent plants. There is also evidence of clustering for seeder 3. The posterior predictive intervals for seeder 5 do not provide evidence against our model.

To look for interactions between the seeder species, we finally consider cross L functions for the 10 pairs of seeders. The lower right posterior predictive plot for seeder 2 vs. seeder 3 in Figure 5 indicates repulsion at small distances and otherwise positive association between these seeders while the remaining plots (not shown) do not contradict the assumptions of independence between the seeders.

[Figure 5 about here.]

5.2.3 Biological interpretation

Ecologists are particularly interested in revealing how individuals interact and whether this interaction varies between species (Uriarte *et al.*, 2004). Our results clearly indicate that resprouter-seeder interaction may be both negative and positive and that the same species can have both positive and negative inter-species interactions.

Some of the interactions revealed here can be explained from ecological background knowledge. For instance, the negative interaction of the most abundant resprouter *Alexgeogia nitens* might be a result of the dense root mat formed by this species making it very difficult for plants from other species to establish themselves close to them. Similarly, the strong positive interactions between the resprouters 15 or 18 and seeder 2 may be explained by specific associations between soil fungi and the plant roots of the seeder. These associations, termed ericoid mycorrhiza, facilitate nutrient uptake from the soil by

plants from the *Ericaceae* family. They enable the seeder to use very complex organic material produced, for example, by resprouters 15 and 18, which is normally impossible to extract from the soil. As a result, individuals from seeder 2 find more nutrients close to resprouters 15 and 18, leading to a positive interaction. As all seeder species in our analysis are from the *Ericaceae* family, other positive interactions between the seeder species and species from the *Banksia* family (resprouters 15 to 18) might be explained in a similar way.

However, for a large number of interactions no biological explanation can be found. There clearly is a need for further biological research to understand these aspects and to yield a better understanding of the overall community dynamics.

6 Concluding remarks

From a statistical point of view, our analysis shows the difficulty of modelling spatial interactions in a plant community which requires very complex models with a large number of parameters. In this situation, we found the Bayesian approach more useful than the frequentist approach as it allowed a more flexible and realistic model. Furthermore, taking biological background information into account in our analysis naturally lead to a hierarchical model. However, we are aware that the model does not sufficiently capture all interactions that may be present in the dataset. It does not consider an intra-species interaction for each seeder type and, similarly, assumes that the seeder species are independent given the resprouters. Incorporating all these aspects into a single model, though, is computationally very hard. Note that ignoring intra-species clustering does not necessarily invalidate estimates of intensity function parameters. Schoenberg (2004) and Waagepetersen (2007), for example, suggest to estimate intensity functions for clustered

point processes using estimating functions given by Poisson likelihood scores. However, it is clear that ignoring clustering leads to too narrow confidence intervals/posterior credibility intervals. Hence the results regarding significant parameters in Sections 4.1 and 5.2.1 should be taken with a pitch of salt.

From an ecological perspective, we are able both to confirm existing knowledge on species' interactions and to generate new biological questions and hypotheses on species' interactions. Finally note that modelling may benefit from including additional information on the individual plants, i.e. a measure of their size or resource availability at the specific location based on the directional distribution of competitors such as described in Miina and Pukkala (2002) if such data were available. In the context discussed here this would require a marked point process modelling approach using an interaction function that takes the properties of each plant and its competitors into account.

Acknowledgement We are grateful to Paul Armstrong for sharing his data with us and for many discussions which provided the biological background for our modelling approach and for the elicitation of informative priors for the interaction radii. The comments from two anonymous referees are gratefully acknowledged too. Supported by the Danish Natural Science Research Council, grant 272-06-0442, 'Point process modelling and statistical inference'.

References

Armstrong, P. (1991). *Species patterning in the heath vegetation of the Northern Sandplain*. Honours thesis, University of Western Australia.

Armstrong, P. (2005). Personal communication.

Baddeley, A., Møller, J. & Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* **54**, 329–350.

Baddeley, A. J. & Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* **12**, 1–42.

Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester.

Beard, J. (1984). Biogeography of the Kwongan. In: *Kwongan – plant life of the sandplains* (eds. J. Pate and J. Beard), University of Western Australia Press, Nedlands, 1–26.

Berman, M. & Turner, R. (1992). Approximating point process likelihoods with GLIM. *Applied Statistics* **41**, 31–38.

Burslem, D., Garwood, N. & Thomas, S. (2001). Tropical forest diversity – the plot thickens. *Science* **291**, 606–607.

Callaway, R. M. (1995). Positive interactions among plants. *The Botanical Review* **61**, 306–349.

Crawley, M. (1997). Biodiversity. In: *Plant Ecology* (ed. M. Crawley), Blackwell Publishing, 325–358.

Dale, M., Dixon, P., Fortin, M., Legendre, P., Myers, D. & Rosenberg, M. (2002). Conceptional and mathematical relationships among methods for spatial analysis. *Ecography* **25**, 558–577.

Diggle, P. (2003). *Statistical Analysis of Spatial Point Patterns*. Oxford University Press, Oxford, 2nd edition.

Dixon, K. (2005). Personal communication.

Durrett, R. & Levin, S. (1998). Spatial aspects of interspecific competition. *Theoretical Population Biology* **53**, 30–43.

Elkington, J. (1991). Report on the vegetation at Cooljarloo W.A. Unpublished report for TI02 Corporation, Ekomin Pty. Ltd., South Perth.

Gelman, A., Meng, X. L. & Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica* **6**, 733–807.

Grabarnik, P. & Särkkä (2006). Modelling of the spatial structure of a forest stand by Gibbs point processes with hierarchical interactions. In preparation.

Gratzer, G., Waagepetersen, R., Splechtna, B. E., Laister, M. & Coomes, D. (2006). The influence of seed dispersal and environmental heterogeneity for generation of spatial patterns of seedlings in a spruce beech fir old growth forest. In preparation.

Greig-Smith, P. (1983). *Quantitative Plant Ecology*. Blackwell Scientific, Oxford.

Herben, T., During, H. & Law, R. (2000). Spatio-temporal patterns in grassland communities. In: *The Geometry of Ecological Interactions: Simplifying Spatial Complexity* (eds. U. Dieckmann, R. Law and J. Metz), Cambridge University Press, Cambridge, 11–27.

Högmander, H. & Särkkä, A. (1999). Multitype spatial point patterns with hierarchical interactions. *Biometrics* **55**, 1051–1058.

Kühlmann-Berenzon, S., Heikkinen, J. & Särkkä, A. (2005). An additive edge correction for the influence potential of trees. *Biometrical Journal* **47**, 517–526.

Kuuluvainen, T. & Pukkala, T. (1889). Effect of scots pine seed trees on the density of ground vegetation and tree seedlings. *Silva Fennica* **23**, 159–167.

Law, R., Herben, T. & Dieckmann, U. (1997). Non-manipulative estimates of competition coefficients in grassland communities. *Ecology* **85**, 505–517.

Law, R., Murrell, D. & Dieckmann, U. (2003). Population growth in space and time: spatial logistic equations. *Ecology* **84**, 252–262.

Liebhold, A. & Gurevitch, J. (2002). Integrating the statistical analysis of spatial data in ecology. *Ecography* **25**, 553–557.

Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J., Hector, A., Hooper, D., Huston, M., Raffaelli, D., Schmid, B., Tilman, D. & Wardle, D. (2001). Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294**, 804–808.

Magurran, A. (1988). *Ecological Diversity and its Measurement*. University Press, Cambridge.

Mateu, J., Us'o, J. & Montes, F. (1998). The spatial pattern of a forest ecosystem. *Ecological Modelling* **108**, 163–174.

Miina, J. & Pukkala (2002). Application of ecological field theory in distance-dependent growth modelling. *Forest Ecology and Management* **161**, 101 – 107.

Møller, J. & Waagepetersen, R. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton.

Murrell, D., Purves, D. & Law, R. (2001). Uniting pattern and process in plant ecology. *Trends in Ecology and Evolution* **16**, 529–530.

Økland, R. H., Rydgren, K. & Økland, T. (1999). Single-tree influence on understorey vegetation in norwegian boreal spruce forest. *Oikos* **87**, 488 – 498.

Richardson, D., Cowling, R., Lamont, B. & van Hensbergen, H. (1995). Coexistence of Banksia species in southwestern Australia: the role of regional and local processes. *Journal of Vegetation Science* **6**, 329–342.

Robert, C. P. & Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer-Verlag, New York.

Schoenberg, F. P. (2004). Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference* **128**, 79–93.

Stoll, P. & Weiner, J. (2000). A neighbourhood view of interactions among individual plants. In: *The Geometry of Ecological Interactions: Simplifying Spatial Complexity* (eds. U. Dieckmann, R. Law and J. Metz), Cambridge University Press, Cambridge, 11–27.

Uriarte, M., Condit, R., Canham, C. & Hubbell, S. (2004). A spatially explicit model of sapling growth in a tropical forest: does indentity of neighbours matter. *Journal of Ecology* **92**, 348–360.

Waagepetersen, R. (2005). Posterior propriety for Poisson processes. Manuscript available at `http://www.math.aau.dk/~rw`.

Waagepetersen, R. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, to appear.

Wiegand, T. & Moloney, K. (2004). Rings, circles, and null-models for point pattern analysis in ecology. *Oikos* **104**, 209–229.

Wu, H., Sharpe, P. J. H., Walker, J. & Penridge, L. K. (1985). Ecological field theory: a spatial analysis of resource interference among plants. *Ecological Modelling* **29**, 215 – 243.

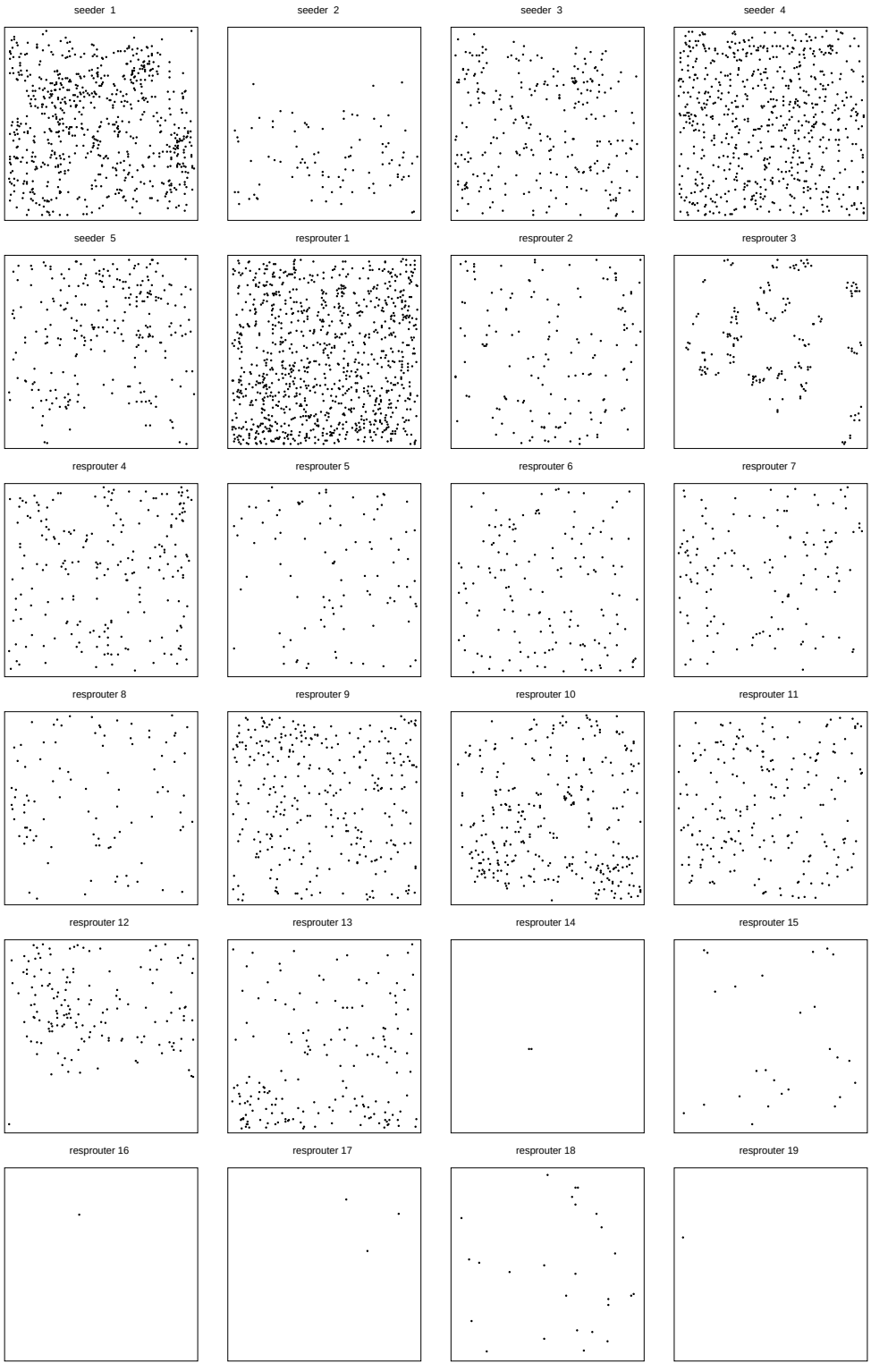

Figure 1: Observed point patterns for the 5 most abundant species of seeders and the 19 most dominant (influential) species of resprouters.

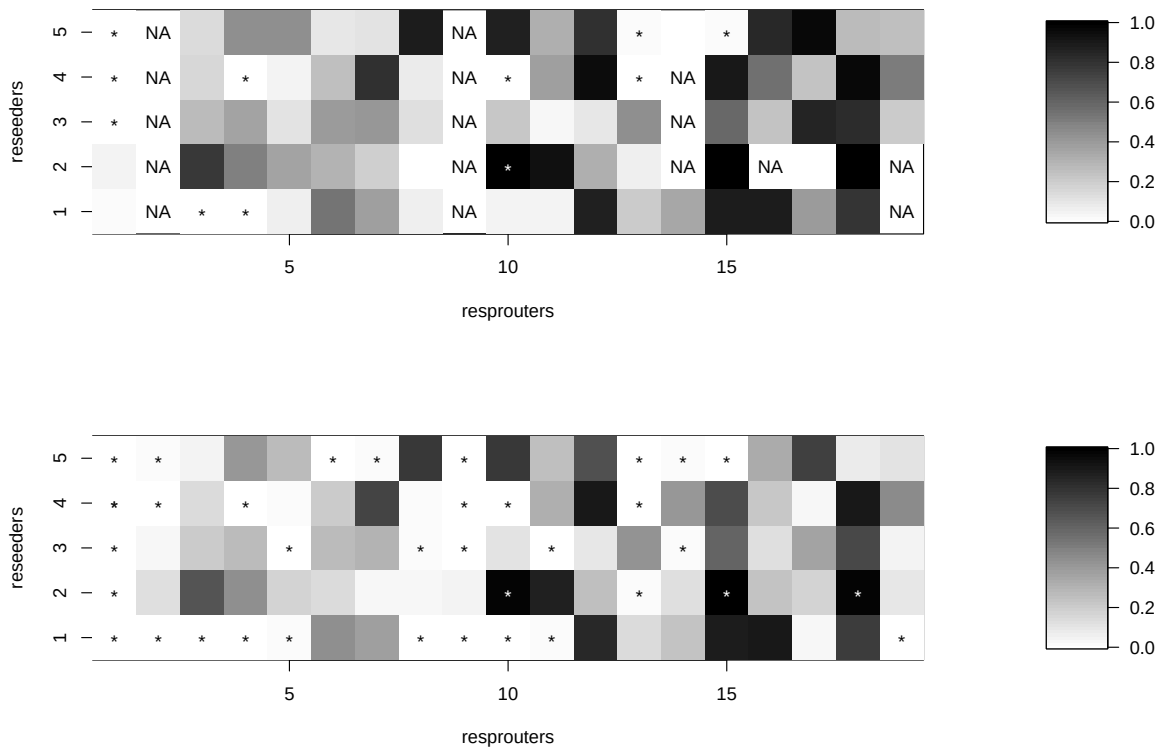

Figure 2: Upper plot: Grey scale plot of $u = \Phi(z)$-statistics for the interaction parameters θ_{ij}. Fields for parameters which are significantly different from zero at the 5 % level are marked with a *. Fields marked with NA correspond to θ_{ij} where the MLE does not exist. Lower plot: Grey scale plot of posterior probabilities $P(\theta_{ij} > 0|\mathbf{y})$. The starred fields are those for which 0 is outside the central 95 % posterior interval for θ_{ij}.

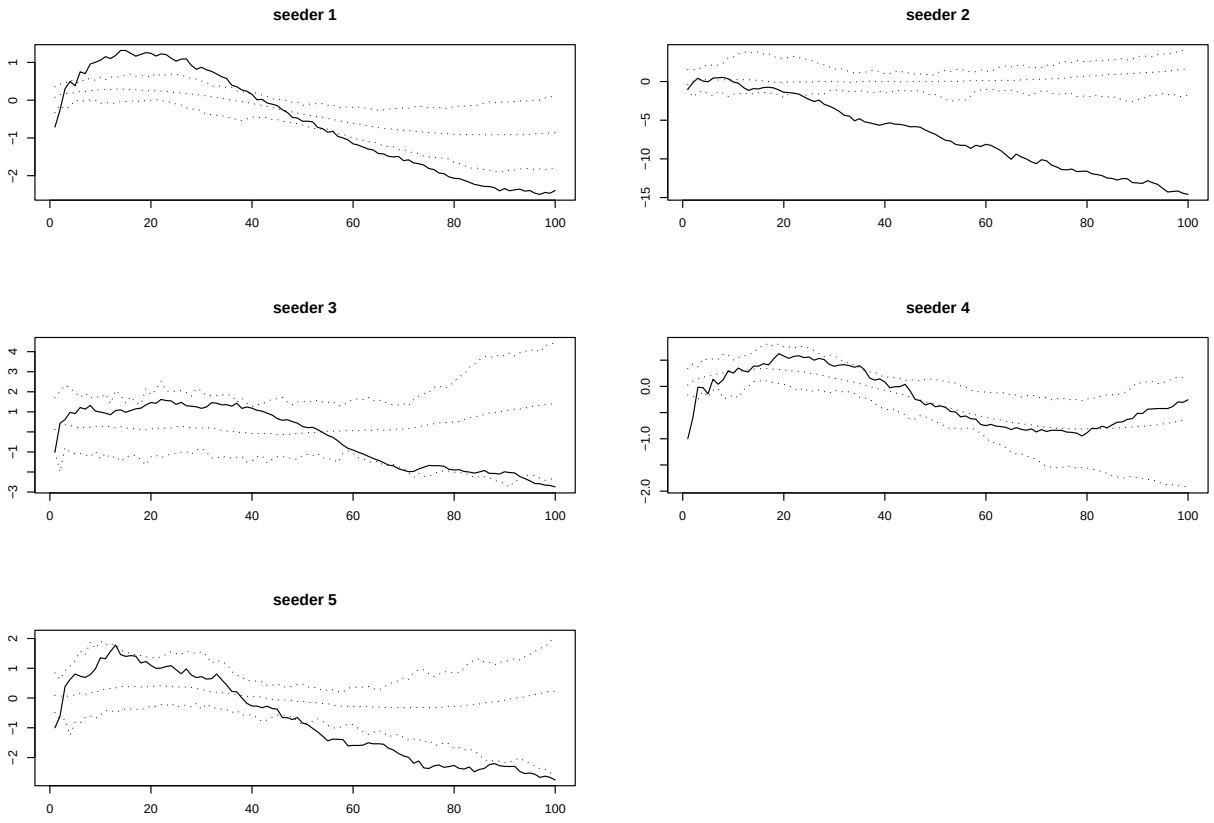

Figure 3: Estimated inhomogeneous $(L(r) - r)$-functions for seeders 1-5 with 95% envelopes simulated from the model. Distance $r > 0$ is in cm.

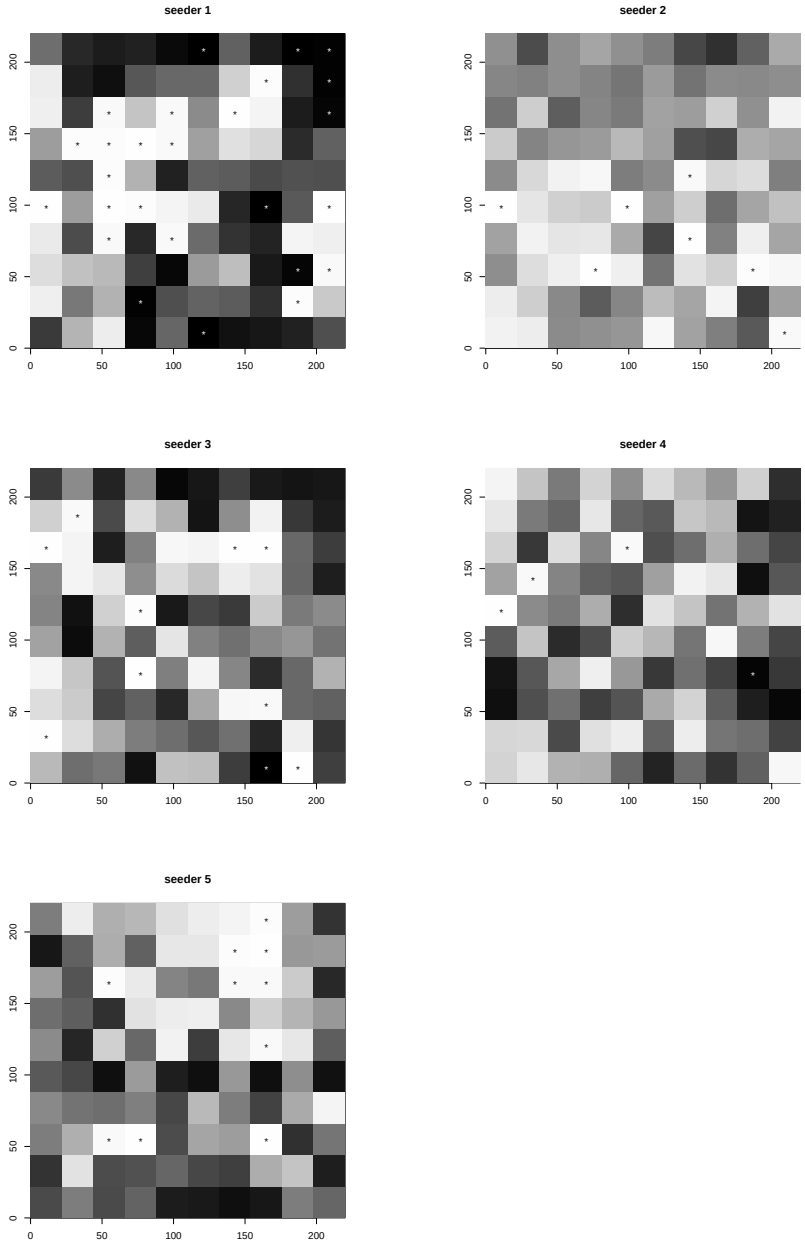

Figure 4: Residual plots based on quadrat counts. Quadrats with a '*' are where the observed counts fall below the 2.5% quantile (white '*') or above the 97.5% quantile (black '*') of the posterior predictive distribution. The grey scales reflect the probabilities that counts drawn from the posterior predictive distribution are less or equal to the observed quadrat counts (dark means small probability).

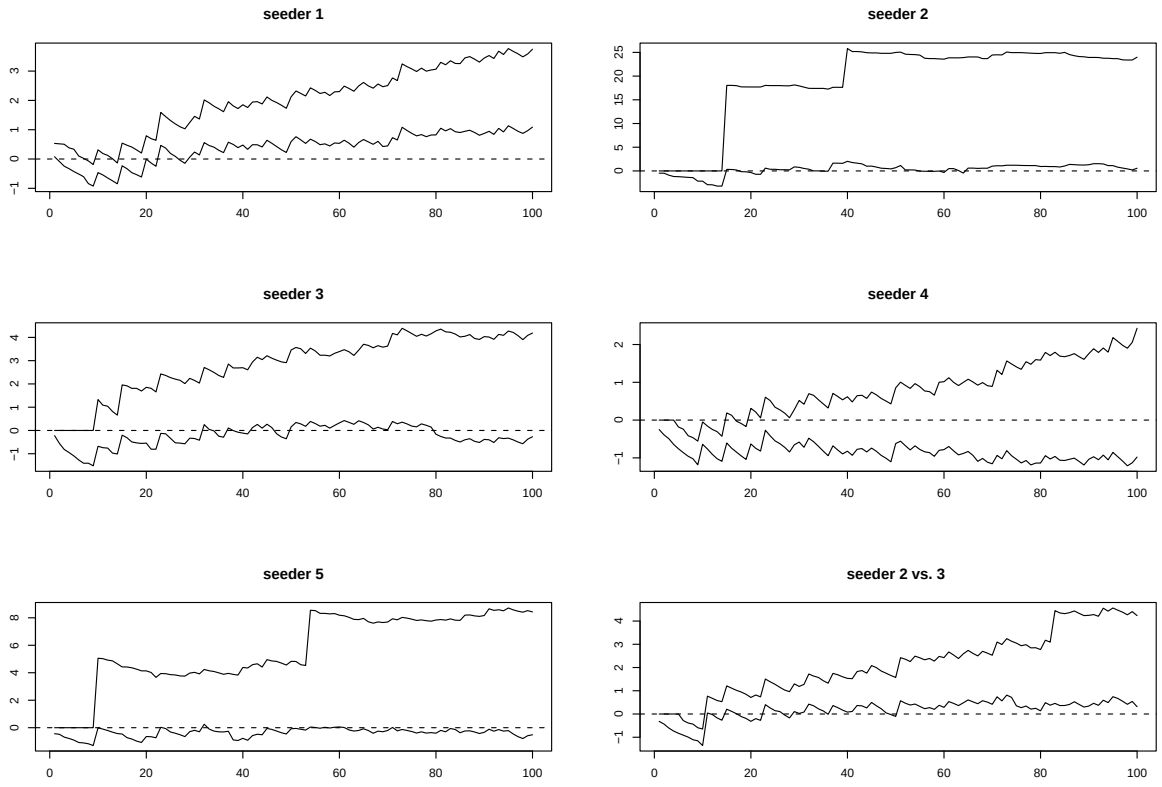

Figure 5: Upper and lower boundaries (solid lines) of the 95% credibility interval for the posterior predictive distribution of $\Delta_i(r)$ for seeders $i = 1, \ldots, 5$. Lower left plot is similar to the other plots but based on the cross L-function between seeder 2 and 3. Distance $r > 0$ is in cm.

1. *Alexgeorgia nitens:*	10-40	11. *Phlebocaria filifolia:*	20-30
2. *Conostylis canescens:*	5-15	12. *Restio sinuousus:*	25-75
3. *Dasypogon bromelifolius:*	15-60	13. *Scholtzia aff. involuc.:*	30-50
4. *Eremae astrocarpa:*	25-75	14. *Allocasuarina humilis:*	50-130
5. *Hibbertia crassifolia:*	10-25	15. *Banksia attenuata:*	150-400
6. *Hibbertia hypericoides:*	10-20	16. *Banksia grandis:*	50-200
7. *Hibbertia subvaginata:*	10-25	17. *Banksia ilicifolia:*	50-200
8. *Hypocalymma xantop.:*	10-25	18. *Banksia menziesii:*	50-250
9. *Lomandra sp.:*	2-10	19. *Eucalyptus todtiana:*	10-250
10. *Lyginia barbata:*	20-100		

Table 1: Range of zone of influence (in cm) for resprouters.