ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Applied Statistics
Series C

Appl. Statist. (2018)
67, *Part* 5, *pp.* 1237–1273

Detecting multivariate interactions in spatial point patterns with Gibbs models and variable selection

T. Rajala, D. J. Murrell and S. C. Olhede

University College London, UK

[Received May 2017. Revised March 2018]

Summary. We propose a method for detecting significant interactions in very large multivariate spatial point patterns. This methodology thus develops high dimensional data understanding in the point process setting. The method is based on modelling the patterns by using a flexible Gibbs point process model to characterize point-to-point interactions at different spatial scales directly. By using the Gibbs framework significant interactions can also be captured at small scales. Subsequently, the Gibbs point process is fitted by using a pseudolikelihood approximation, and we select significant interactions automatically by using the group lasso penalty with this likelihood approximation. Thus we estimate the multivariate interactions stably even in this setting. We demonstrate the feasibility of the method with a simulation study and show its power by applying it to a large and complex rainforest plant population data set of 83 species.

Keywords: Barro Colorado Island; Gibbs models; Multivariate point patterns; Species interaction; Variable selection

1. Introduction

Spatial point patterns are a common form of observation in plant ecology (Waagepetersen *et al.*, 2016), epidemiology (Diggle *et al.*, 2005), astrophysics (Stoica *et al.*, 2007), seismology (Schoenberg, 2003), social science (Amburgey, 1986), medicine (Olsbo *et al.*, 2013) and criminology (Mohler *et al.*, 2011). Although understanding single, univariate spatial point patterns and their generating point processes is important, frequently we observe *labelled* point processes or, more precisely, multiple types of points. The prevalence of multivariate point processes is particularly noticeable in plant ecology where there may be many tens or hundreds of types (species) (Flügge *et al.*, 2014; Baldeck *et al.*, 2013a, b; Kanagaraj *et al.*, 2011; Punchi-Manage *et al.*, 2013). Such processes have seen much less study in the statistical literature than univariate processes and present some novel challenges, as we shall explain and address in this paper.

To be able to make sense of multivariate point processes, we focus on addressing three important outstanding problems in understanding interactions:

(a) characterizing patterns that are associated with both small and large scales simultaneously,
(b) characterizing multiple features spanning more than one variable and
(c) estimating such patterns stably.

Address for correspondence: T. Rajala, Department of Statistical Science, University College London, Gower Street, London, WC1E 6BT, UK.
E-mail: t.rajala@ucl.ac.uk

The effect of including these characteristics in our studies of a multivariate point process greatly increases the number of potential parameters to characterize and naturally leads to difficulties in arriving at unique and stable parameter estimates.

Early approaches to modelling and analysing more than one point process focused on bivariate representation and analysis (Diggle and Milne, 1983; Amburgey, 1986; Brix and Møller, 2001; Gelfand *et al*., 2004; Shimatani, 2001; Diggle *et al*., 2005), and going much beyond the bivariate case has proved very challenging. To the best of our knowledge the most diverse model-based analysis of multivariate point patterns to date is of nine rainforest tree species by Waagepetersen *et al*. (2016), who used a multivariate log-Gaussian Cox process model, and fitted the model by using non-parametric least squares. The reason given for confining studies to only nine species by them was to limit the computational burden of analysis. This work is inspirational, but our motivating problem in this paper is to analyse jointly an order of magnitude more species. More specifically, there are 300 species (types) in the full data set that was used in Waagepetersen *et al*. (2016), and we wish to extend analysis to investigate as many of these species as possible. This brings us into the realm of high dimensional statistics as the number of interactions scales exponentially in the number of species, whereas the number of points scales only linearly in the number of species. To deal with this inconvenient scaling we shall need to use shrinkage, as is commonly done in high dimensional data analysis, and has already been developed for regression problems and covariance estimation; see for example van der Geer and Bühlmann (2011). In the context of point processes, because estimation is not implemented with linear methods, penalization needs to be deployed carefully. Starting from ideas of Baddeley *et al*. (2014) we shall use a generalized regression-based fitting approach, and so we borrow ideas from second-generation penalized regression rather than from matrix shrinkage, even if we are estimating co-associations rather than a mean intensity.

A second significant problem in point process modelling is proposing models, and associated estimation methods, that yield sufficient multiscale behaviour, e.g. variability at fine scales, as well as over medium to long scales. It is all very well to posit variability for fine scales by using a log-Gaussian Cox process but, as estimation is normally based on some form of averaging, unless the random intensity is very high locally, it will be impossible to estimate the log-Gaussian Cox process's generating mechanism as we shall not have enough points. For longer spatial scales the log-Gaussian Cox process is a well-suited modelling framework, but it is not a good framework for studying small-scale interactions. Instead we shall use the multivariate Gibbs point process model to discover small-scale point-to-point interactions in the same Barro Colorado Island (BCI) rainforest data set that was studied by Waagepetersen *et al*. (2016).

Key to our modelling and estimation is therefore capturing an appropriate degree of sample heterogeneity. The estimation framework that we shall introduce can take into account variations that are associated with

(a) habitat associations, i.e. correlation of species presence in the landscape with known environmental covariates, with
(b) dispersal mechanisms and competition such as seedling clustering or self-thinning, and with
(c) attraction and repulsion between the small-scale locations between different species.

To demonstrate utility, we shall fit the model to an adult plant community consisting of 83 species, which is an order of magnitude more species than Waagepetersen *et al*. (2016) analysed.

The analysis of multivariate point patterns of more than a handful of species (usually two, e.g. Brix and Møller (2001), Diggle and Milne (1983), Högmander and Särkkä (1999) and Funwi-Gabga and Mateu (2012)) has mostly relied on non-parametric estimation techniques.

Many ecological analyses of large and diverse rainforest data sets have been carried out by using the K-function or similar non-parametric summaries, either directly by comparing the summary values under various null model scenarios, or indirectly as part of minimum contrast model fitting of Cox processes (e.g. Lan *et al*. (2012), Flügge *et al*. (2014), Waagepetersen *et al*. (2016), Yang *et al*. (2016), Velázquez *et al*. (2016) and Brown *et al*. (2016)). In the Gibbs model framework, the three-variate analysis of different size trees by Grabarnik and Särkkä (2009) is perhaps the most extensive modelling approach and closest in spirit to the work of this paper. We shall extend their trivariate case to a full multiscale multivariate interaction Gibbs model.

Direct likelihood inference for Gibbs models is not available because of intractable normalizing constants, but several pseudolikelihood approximations are available, such as the 'Berman–Turner machine' (Baddeley and Turner, 2000) that was used by Grabarnik and Särkkä (2009) which casts the pseudolikelihood estimation equation as a Poisson regression problem and subsequently estimates the model with standard statistical software. The Poisson regression approach is surpassed in accuracy by the logistic regression approach that was developed by Baddeley *et al*. (2014), who formulated the pseudolikelihood estimation equation as a logistic regression using auxiliary dummy point configurations. Again, very conveniently, standard statistical software can be used to fit the model.

With a model and the likelihood approximation at hand, we shall next tackle the issue of high dimensional variable selection: any reasonable model for a p-variate point pattern will have a high number of parameters when $p \gg 3$, scaling at least like $\mathcal{O}(p^2)$. As an illustration, the model that we present estimates intraspecies interactions and pairwise species-to-species interactions, on three different spatial scales, and includes six covariates, giving a total of about 11 000 parameters, to be compared with the number of observations of 31 650 points, making the number of parameters and observations of the same order. To discover significant interactions in such a high dimensional setting, we shall use recent research on penalized optimization. Several techniques are available, and to show proof of concept we shall be using the group lasso (Yuan and Lin, 2006; Meier *et al*., 2008). We shall also perform a limited comparison with the Bayesian spike-and-slab variable selection approach (Mitchell and Beauchamp, 1988). Using penalized regression (the group lasso) rather than full posterior inference to compute a maximum *a posteriori* estimate yields both simplicity in interpretation and computational speed. This enables us to study several 'priors', or penalization choices, at once and thus this enables us to make fewer assumptions on the generating mechanism of the data, in fact enabling us to explore the properties of our modelling framework.

We start in Section 2 by introducing the Gibbs model; then we recall the chosen likelihood approximation and discuss the chosen variable selection techniques in detail. In Section 3 we recall for comparison the non-parametric Monte Carlo (MC) technique that is often used for analysing p-variate patterns when $p \gg 1$. In Section 4 we test the method on several increasingly complex simulation scenarios, to obtain a better understanding of the performance of the method. In Section 5 we apply the method to the BCI rainforest data (Condit, 1998) that inspired these developments. We conclude in Section 6, and discuss outstanding problems and future avenues of investigation.

2. The Gibbs model and Gibbs model fitting with variable selection

Let the observed multivariate point pattern be a set of labelled point locations $\mathbf{x} = \{(x, t)\}$ where $x \in W$ are the observed locations inside a known, bounded observation window $W \subset \mathbb{R}^d$, $d = 1, 2, 3, \ldots$ (for the problems that we shall study, $d = 2$), and $t \in \{1, \ldots, p\}$, $p \geqslant 1$, are categorical labels (types or, in our case, species) attached to each location. Denote the type i subpattern by

$\mathbf{x}_i = \{x : (x, i) \in \mathbf{x}\}$. Write $n_i(A) = \#(\mathbf{x}_i \cap A)$ for the count of points of type i in a set $A \subset \mathbb{R}^d$, and write $n_i = n_i(W)$. As is standard let $b(u, r)$ denote a ball of radius r centred at $u \in \mathbb{R}^d$.

2.1. Gibbs models for point patterns

Our model for multivariate spatial point patterns is part of the Gibbs point process model family; we refer to van Lieshout (2000), who detailed general properties of Markov point processes (where Gibbs point processes are a special case), and Illian *et al.* (2008) and Chiu *et al.* (2013), who discussed general point processes. We shall use a Gibbs model, defined from a set of *potential functions* ϕ_{ij}, to have a probability density of the form

$$f(\mathbf{x}) \propto \exp\left\{ \sum_{i,j=1}^{p} \phi_{ij}(\mathbf{x}_i, \mathbf{x}_j) \right\}, \tag{1}$$

with respect to the unit rate Poisson process μ_1 on W. The normalizing constant for the density,

$$\int_{\mathcal{X}} \exp\left\{ \sum_{i,j=1}^{p} \phi_{ij}(\mathbf{x}_i, \mathbf{x}_j) \right\} \mu_1(\mathrm{d}\mathbf{x}),$$

with $\mathcal{X}$ the space of all locally finite multitype point patterns, is in practice intractable for all except the homogeneous Poisson models where $\phi_{ij} \equiv \alpha_i \in \mathbb{R}$. The functions ϕ_{ij} are used to specify the model class member. We shall use the special form of

$$\phi_{ij}(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} \sum_{x \in \mathbf{x}_i} \alpha_i^{\mathrm{T}} \mathbf{z}_i(x) + \sum_{x \in \mathbf{x}_i} \beta_{ii}^{\mathrm{T}} \mathbf{g}_{ii}(x, \mathbf{x}_i \setminus x), & i = j, \\ \sum_{x \in \mathbf{x}_i} \beta_{ij}^{\mathrm{T}} \mathbf{g}_{ij}(x, \mathbf{x}_j), & i \neq j. \end{cases} \tag{2}$$

In expression (2) the parameters $\alpha_i \in \mathbb{R}^{K_i+1}$ regulate intensity and covariate effects (so-called first-order effects), and $\beta_{ij} \in \mathbb{R}^{K_{ij}}$ are parameters for interactions between the locations (second-order effects). The β_{ij}s give the magnitudes of the interactions, whereas the vector-valued functions $\mathbf{g}_{ij}$, which we specify later in Section 2.2, determine the form, spatial scales and orders of the interactions. The covariate effects $\mathbf{z}_i(u) = (1 \; z_{i1}(u) \; z_{i2}(u) \ldots z_{iK_i}(u))^{\mathrm{T}}$ represent the baseline effect and any covariate and trend effect values that we have at locations $u \in W$. In this formulation we assume that the covariates are available everywhere in the window W. This is usually achieved by interpolation from prior data collection efforts.

To obtain a heuristic understanding of the model, assume first that all parameters except $\alpha_i \in \mathbb{R}$ are 0 and $\mathbf{z}_i \equiv 1$. Then density (1) becomes $\exp(\sum_{i=1}^{p} n_i \alpha_i) = \prod_{i=1}^{p} \exp(n_i \alpha_i)$, which is the likelihood of a collection of p independent homogeneous Poisson processes. Now add some non-constant covariates $z_{i2}(u), z_{i3}(u), \ldots$ and set $\alpha_{i2}, \alpha_{i3}, \ldots \neq 0$: the model becomes a collection of independent and inhomogeneous Poisson processes. We subsequently add intratype interaction terms by letting $\beta_{ii} \neq 0$: the independent components are no longer Poisson but exhibit internal, within-type point-to-point interactions (attraction or repulsion depending on the sign of β_{ii}). Finally we can add intertype interaction terms by setting $\beta_{ij} \neq 0$: the locations of different types are no longer independent.

Model (1) is log-linear in parameters α and β. We collect the covariate parameters in the vector $\theta_0 = (\alpha_1^{\mathrm{T}} \ldots \alpha_p^{\mathrm{T}})^{\mathrm{T}}$, the intratype interaction parameters to $\theta_1 = (\beta_{11}^{\mathrm{T}} \ldots \beta_{pp}^{\mathrm{T}})^{\mathrm{T}}$ and intertype interaction parameters, assuming for now that i and j interact symmetrically, to $\theta_2 = (\beta_{12}^{\mathrm{T}} \ldots \beta_{(p-1)p}^{\mathrm{T}})^{\mathrm{T}}$. Subsequently we collect them in the vector $\theta = (\theta_0^{\mathrm{T}} \; \theta_1^{\mathrm{T}} \; \theta_2^{\mathrm{T}})^{\mathrm{T}}$. Then the density in model (1) can be written in the form

$$f(\mathbf{x}) = f_\theta(\mathbf{x}) \propto \exp(\theta^{\mathrm{T}} \mathbf{v}), \tag{3}$$

where the vector $\mathbf{v} = \mathbf{v}(\mathbf{x}) = (\mathbf{s}_0^{\mathrm{T}} \; \mathbf{s}_1^{\mathrm{T}} \; \mathbf{s}_2^{\mathrm{T}})^{\mathrm{T}}$ has components

(a) $\mathbf{s}_0 = (\mathbf{s}_1^{\mathrm{T}} \dots \mathbf{s}_p^{\mathrm{T}})^{\mathrm{T}}$ with $\mathbf{s}_i = \Sigma_{x \in \mathbf{x}_i} \mathbf{z}_i(x)$,
(b) $\mathbf{s}_1 = (\mathbf{s}_{11}^{\mathrm{T}} \dots \mathbf{s}_{pp}^{\mathrm{T}})^{\mathrm{T}}$ with $\mathbf{s}_{ii} = \Sigma_{x \in \mathbf{x}_i} \mathbf{g}_{ii}(x, \mathbf{x}_i \setminus x)$ and
(c) $\mathbf{s}_2 = (\mathbf{s}_{12}^{\mathrm{T}} \dots \mathbf{s}_{(p-1)p}^{\mathrm{T}})^{\mathrm{T}}$ with $\mathbf{s}_{ij} = \Sigma_{x \in \mathbf{x}_i} \mathbf{g}_{ij}(x, \mathbf{x}_j)$, $i < j$.

The vector $\mathbf{v} = \mathbf{v}(\mathbf{x})$ is written as a matrix ($\mathbf{v}^{\mathrm{T}}$) and takes the role of a design matrix in the standard regression setting. The matrix has only one row when the point process is observed only once. Independent replicates would result in additional rows in the matrix.

Given data $\mathbf{x}$, covariates $\mathbf{z}$ and the interaction functions $\mathbf{g}_{ij}$, the vector $\mathbf{v}$ is fixed and the non-normalized model (3) is log-linear in the unknown coefficients θ. The model therefore belongs to the family of exponential Gibbs models, and we can apply inference techniques that are designed for the exponential Gibbs family.

2.2. The interaction functions

We now define the exact model that we shall use in our examples. Several definitions are available for the interaction functions $\mathbf{g}_{ij}$ in expression (2). We assume that $\mathbf{g}_{ij}$ are non-negative functions to remove sign ambiguity when estimating β_{ij}. The most popular class of models is the pairwise interacting models with

$$\mathbf{g}_{ij}(x, \mathbf{x} \setminus x) = \sum_{y \in \mathbf{x} \setminus x} \psi_{ij}(\|x - y\|),$$

for some functions $\psi_{ij} : \mathbb{R}_+ \mapsto \mathbb{R}_+^{K_{ij}}$. The most common choice in the case of $K_{ij} = 1$ is the Strauss model

$$g(x, \mathbf{x} \setminus x) = \sum_{y \in \mathbf{x} \setminus x} \mathbf{1}(\|x - y\| < r), \qquad r > 0, \tag{4}$$

effectively counting the number of r-close pairs in the pattern. In the univariate case the Strauss model is valid only when the corresponding interaction coefficient $\beta < 0$ so fewer point pairs lead to a higher likelihood. Trying to model positive interactions, or clustering, leads to an unstable model that produces patterns of singular megaclusters, so the case $\beta > 0$ is excluded for the simple Strauss model (Gates and Westcott, 1986).

To circumvent this limitation Geyer (1999) introduced a model which he called the saturation model that still defines a locally stable process even with a positive interaction parameter $\beta > 0$. In the simple univariate case the saturation model is defined via

$$g(x, \mathbf{x} \setminus x) = \min\{c, \#[(\mathbf{x} \setminus x) \cap b(x, r)]\}, \qquad r > 0, \quad c \in \mathbb{N}_+, \tag{5}$$

where the range r is the reach of the Euclidean neighbourhood, and c is a saturation level. In this model, each point contributes to the likelihood a factor that is relative to the number of r-neighbours or c, whichever is smaller (hence the saturation). In ecological terms, the Geyer model can capture the fact that individuals may cluster at some distances but are likely to segregate at shorter distances because of intense competition, and the saturation parameter reproduces the feature that the neighbourhood must eventually saturate with individuals as resources are finite. The model belongs to a class of interacting neighbour models (Grabarnik and Särkkä, 2001), so named because the conditional intensity (8) for this model depends not just on the local neighbourhood of a point in u (which it does for pairwise models), but also on the neighbourhoods of the neighbours of u.

To model several types of points and more than one spatial scale, we generalize the models in two ways by adding

(a) multiple ranges and

(b) cross-type interactions when $p \geqslant 2$.

Let $\mathbf{r}_{ij} = \{r_{ijk} : 0 = r_{ij0} < r_{ij1} < \ldots < r_{ijK_{ij}}\}$ be a fixed increasing vector of ranges for $i \leqslant j$. Let $c_{ij} = \{c_{ijk} \in \mathbb{N}\}$ be the saturation parameters. For a given range vector, write

$$\mathrm{ne}_{ijk}(x, \mathbf{x}) := \#\{\mathbf{x} \cap [b(x, r_{ijk}) \setminus b(x, r_{ij(k-1)})]\},$$

for the number of neighbours of $\mathbf{x}$ that a point x has in the annulus between ranges $r_{ij(k-1)}$ and r_{ijk}. Then the multistep multitype Strauss model is defined by using the interaction function made of components

$$g_{ijk}(x, \mathbf{x}) := \mathrm{ne}_{ijk}(x, \mathbf{x}), \tag{6}$$

and the multistep multitype saturation model of components

$$g_{ijk}(x, \mathbf{x}) := \min\{c_{ijk}, \mathrm{ne}_{ijk}(x, \mathbf{x})\} \tag{7}$$

in definitions of $\mathbf{g}_{ij}$. Note that the Strauss model is a special case of the saturation model with $c \to \infty$.

Several other forms of g_{ijk} can be used, and multiple forms can be combined as described in Baddeley *et al*. (2013). We shall not pursue them here as either

(a) they can be approximated by the Strauss or saturation model as the interaction functions $\beta_{ij}^{\mathrm{T}} \mathbf{g}_{ij}$ are step functions over spatial scales,
(b) there is not sufficient data available to estimate very fine details over many spatial scales or
(c) they are computationally costly (e.g. morphological functions).

However, if forms such as the area interaction model (Baddeley and van Lieshout, 1995) seem more appropriate for a specific application, the framework proposed is still valid and can be adapted to be used in this setting.

For fitting the model (Section 2.3), we need to define the conditional (or Papangelou) intensity of the model: at any point $u = (x, i) \in \mathbb{R}^d \times \{1, \ldots, p\}$ let

$$\lambda_\theta(u; \mathbf{x}) := \frac{f_\theta(\mathbf{x} \cup u)}{f_\theta(\mathbf{x} \setminus u)} \tag{8}$$

with $0/0 := 0$. Heuristically, $\lambda_\theta(u; \mathbf{x})\mathrm{d}u$ can be understood as the conditional probability of observing a point u, given the rest of the pattern $\mathbf{x}$. For the exponential family Gibbs models the conditional intensity has the quite simple form of

$$\lambda_\theta(u; \mathbf{x}) = \exp[\theta^{\mathrm{T}}\{\mathbf{v}(x \cup u) - \mathbf{v}(\mathbf{x} \setminus u)\}] = \exp\{\theta^{\mathrm{T}} \mathbf{v}(u; \mathbf{x})\}, \tag{9}$$

where we use the notation $\mathbf{v}(u; \mathbf{x}) = \mathbf{v}(\mathbf{x} \cup u) - \mathbf{v}(\mathbf{x} \setminus u)$. Note that the intractable normalizing constant cancels out in this expression.

The conditional intensity (8) with the stepwise components (6) and (7) is at any marked point $u = (x, i)$

$$\log\{\lambda_\theta(u; \mathbf{x})\} = z(u)^{\mathrm{T}} \alpha_i + \sum_{i=1}^{p} \sum_{j=i}^{p} \sum_{k=1}^{K_{ij}} \beta_{ijk}\, \omega_{ijk}(u, \mathbf{x}_j), \tag{10}$$

with

$$\omega_{ijk}\{(x, i), \mathbf{x}_j\} = g_{ijk}(x, \mathbf{x}_i \setminus x) + \sum_{y \in \mathbf{x}_j} \{g_{ijk}(y, \mathbf{x}_i \cup x) - g_{ijk}(y, \mathbf{x}_i \setminus x)\}.$$

The functions $\omega_{ijk} \in [-c_{ijk}, n_i c_{ijk}]$ track the changes in the neighbourhood inclusion counts within and between different types. An illustration of the model is given in Appendix C.

For both models the non-canonical parameters r_{ijk} and c_{ijk} cannot be directly inferred by the pseudolikelihood methods. Thus the parameters need to be fixed as part of the model definition. In data analysis we usually have *a priori* information on relevant ranges $\{r_{ijk}\}$ (e.g. Uriarte *et al*. (2004)), but the saturation level $\{c_{ijk}\}$ is more difficult to set. To reduce this complexity we propose to choose the saturation levels $\{c_{ijk}\}$ automatically depending on the abundances n_j. The idea is that if the abundance n_j is high the neighbourhoods ne_{ijk} and therefore ω_{ijk} saturate often even under the independence assumption. Under the full independence assumption, the expectation of ω_{ijk} is a function of $c = c_{ijk}$,

$$t(c) := c\{1 - F_a(c-1)\} + a\{F_a(c-1) + F_a(c-2)\},$$

where $a = a_{ijk} = |b(o, r_{ijk}) \setminus b(o, r_{ij(k-1)})| n_j / |W|$ and F_a is the cumulative distribution function of a Poisson(a) random variable (see Appendix B). The function t is non-decreasing and $t \to 2a$ when $c \to \infty$, so $1 - t(c)/(2a)$ is the cumulative distribution function of the event that saturation occurs under independence. To avoid saturation due to high abundances alone, a sensible choice of c is a value for which saturation under independence is unlikely. Therefore, we set a small $0 < \epsilon < 1$ such that

$$1 - t(c)/(2a) < \epsilon$$

and, by further using the approximation $t(c)/(2a) \approx F_a(c-1)$, we choose $c = c_{ijk}$ to be the $(1-\epsilon)$-quantile of F_a. In the examples we use $\epsilon = 0.01$. With the modification the interaction functions $\mathbf{g}_{ij}$ become asymmetric, but we shall treat the β_{ij}s symmetrically in our examples for simplicity.

2.3. *Inference: approximating the likelihood at its mode*

The likelihood in equation (1) (or equivalently equation (3)) is not computationally tractable because of an unspecified normalizing constant. To carry out standard likelihood inference the constant can be approximated by MC techniques, but these tend to be computationally costly even for the univariate case. The more commonly used approach is to use pseudolikelihood techniques, which replace the function to maximize with something that approximates the likelihood at its mode. We shall use the recent developments that were proposed by Baddeley *et al*. (2014), which in practice conveniently lead to a logistic regression formulation.

We have summarized the details of the method in Appendix A. For the discussion, it suffices to know that the pseudolikelihood function $\tilde{f}_\theta$ is formally a likelihood of a logistic regression function. It involves additional sets of random dummy points per type, of which construction, particularly their intensity ρ_i, is an additional user decision. Baddeley *et al*. (2014) discussed several potential options to be used for the dummy distributions and intensities. We shall use the recommended homogeneous stratified uniform distributions and, if not otherwise stated, intensities that are four times the intensity of data.

To address boundary effects due to censoring near the edges of observation window W, we shall exclude components of $\tilde{f}_\theta$, say $\mathbf{b}(u)$, for which the distance from u to the border of W is less than a range $r_{\mathrm{bor}} > 0$, i.e. in the sum in Appendix A, equation (15), the W is replaced by $W \ominus b(o, r_{\mathrm{bor}})$. The range r_{bor} is taken to be the maximal interaction range in the model and is determined by the $\mathbf{g}_{ij}$s.

2.4. Penalized inference for grouped coefficients

Our main goal is the detection and estimation of significant within-type and between-type interactions in multivariate point pattern data. For this the previous section enables us to estimate groups of coefficients, such as the intertype interaction vectors $\beta_{ij} \in \mathbb{R}^{K_{ij}}$ for type pairs $i \neq j$. Subsequently to that, we need to determine whether $\beta_{ij} = 0$ or not.

In a general setting, let the length of an unknown coefficient vector θ be M, and split θ into m smaller groups by using a partition given by $\{\pi_1, \ldots, \pi_m\}$ of $\{1, \ldots, M\}$. For shorthand we shall write $(g) = \pi_g$, where $g = 1, \ldots, m$. Write $e_g = 0$ when $\theta_{(g)} = 0$ and $e_g = 1$ when at least one of the coefficients $\theta_{(g)}$ is non-zero. In our case the partition is given by the types and type pairs, and the e_gs are connected to the events $\beta_{ij} \neq 0$. Our task is to determine which e_gs are non-zero in a given data $\mathbf{x}$, which is a task known as grouped variable selection (Yuan and Lin, 2006).

Variable selection and shrinkage in high dimension, which are usually implemented by using penalized optimization, have become an important computational statistical technique because of an increase in high data throughput applications such as genomics and text analytics. Variable selection has already been applied to point pattern analysis; see for example the work by Yue and Loh (2014), who applied variable selection in the univariate case for covariate selection. Several methods exist for the particular problem of group level selection (or penalization) that we are addressing in this context (see for example Breheny and Huang (2009)). From the available selection of optimization criteria, we picked the group lasso for logistic regression (Meier *et al.*, 2008), as a suitable penalized version of Baddeley *et al.* (2014). The group lasso is an extension of the original l_1-penalization by Tibshirani (1996) for individual coefficients and has been further extended to a mixed level penalization by Simon *et al.* (2013). To be concrete, the group lasso estimator is defined by

$$\hat{\theta}_\gamma = \arg\max_\theta L_\gamma(\mathbf{x}; \theta),$$

with the group penalized likelihood function

$$L_\gamma(\mathbf{x}; \theta) := \log\{\tilde{f}_\theta(\mathbf{x})\} - \gamma \sum_{g=1}^{m} |(g)|^{-1/2} \|\theta_{(g)}\|_2, \tag{11}$$

where $\gamma > 0$ is a penalization parameter and $\|\cdot\|_2$ is the Euclidean distance. Some groups $\theta_{(g)}$ are shrunk to exactly 0, depending on the strength of the penalization, so the group lasso does provide us with binary indicators $e_g = \mathbf{1}(\hat{\theta}_{(g)} \neq 0)$.

The penalization parameter γ is directly linked to the amount of non-zero e_gs that the algorithm outputs and it needs to be chosen by the user. With some extra computational cost, we compute what is known as the lasso path: a set of estimates for each value of $\gamma \in \Gamma = [0, \gamma_{\max}]$. The maximal penalty $\gamma_{\max}$ is the penalty level below which the first penalized group is let into the model. In practice, we shall use by default a 100-step log-linearly decreasing equidistant grid from $\gamma_{\max}$ to $0.001\gamma_{\max}$.

Typically when using a lasso-based variable selection analysis one would use cross-validation (CV) to choose the level of penalization that gives the best balance between model fit and quality of prediction, or minimizes the expected risk (Giraud (2014), section 5.2), also called the extra-sample error (Hastie *et al.* (2001), section 7.1). In the point pattern context, Yue and Loh (2014) also chose their penalty by using CV, but unfortunately no details were given how the data were partitioned or what error metric was used for the prediction. Conducting CV by splitting the constructed data frame that was used in the logistic regression part of the inference is not justified as the rows are dependent because of spatial correlation. This flaw leads us to overestimate the complexity of the model. Note also that it has been observed that, as CV is

aimed at prediction, rather than model selection, too many variables are often retained (van der Geer and Bühlmann (2011), section 2.5.1).

Let us define a CV procedure for spatial point pattern models that coincides with the CV for classical regression. Two elements are needed; a way to split the data into meaningful subsets of independent data, and a way to measure prediction error, or risk, with the already explained *caveats*. Let the observation window W be partitioned into disjoint quadrats $W_1, \ldots, W_K$, and write for each $k = 1, \ldots, K$

$$\hat{\theta}_{\gamma,-k} = \arg\max_{\theta} L_\gamma(\mathbf{x} \setminus W_k; \theta)$$

for the group lasso estimate using all the data *except* points inside W_k. For the CV risk we need to define prediction quality, e.g. the residual sum of squares that are used in linear regression. Baddeley *et al*. (2004) introduced the h-residual measure for point patterns which we can use to define a residual sum of squares. Define the CV h-residual as

$$\begin{aligned}\hat{R}_{\gamma,k} &:= R(W_k, \hat{\theta}_{\gamma,-k}) \\ &= \sum_{u \in \mathbf{x} \cap W_k} h(u, \mathbf{x} \setminus u, \hat{\theta}_{\gamma,-k}) - \int_{W_k \times \{1,\ldots,p\}} h(u, \mathbf{x}, \hat{\theta}_{\gamma,-k}) \lambda_{\hat{\theta}_{\gamma,-k}}(u; \mathbf{x}) \mathrm{d}u \\ &= \sum_{i=1}^{p} \left[\sum_{x \in \mathbf{x}_i \cap W_k} h\{(x,i), \mathbf{x} \setminus (x,i), \hat{\theta}_{\gamma,-k}\} - \int_{W_k} h\{(x,i), \mathbf{x}, \hat{\theta}_{\gamma,-k}\} \lambda_{\hat{\theta}_{\gamma,-k}}\{(x,i); \mathbf{x}\} \mathrm{d}x \right],\end{aligned} \tag{12}$$

where h is a non-negative function called the test function. Baddeley *et al*. (2004) and Coeurjolly and Lavancier (2013) listed four options for the test function, of which we shall look at three. The *raw* residuals with $h = 1$ assess only the trend part of the model, and the *inverse* residuals with $h(u; \mathbf{x}, \theta) = \lambda_\theta(u; \mathbf{x})^{-1}$ assess the interactions. The *Pearson* residuals with $h(u; \mathbf{x}, \theta) = \lambda_\theta(u; \mathbf{x})^{-1/2}$ are similar to the inverse but with variance that is in theory independent of λ for the Poisson process.

We define the K-fold CV estimator of the prediction risk as the mean of the squared residuals

$$\hat{R}_{\mathrm{CV}}(\gamma) := \frac{1}{K} \sum_{k=1}^{K} \hat{R}_{\gamma,k}^2$$

and we choose the minimizer of the estimated risk,

$$\hat{\gamma}_{\mathrm{CV}} := \arg\min_{\gamma \in \Gamma} \hat{R}_{\mathrm{CV}}(\gamma),$$

as the cross-validated penalty level. Note that one can easily weight the residuals by relative importance of each type or the quadrat size.

Fig. 1 illustrates this approach for an example pattern generated by experiment 1 in Section 4.1. The individual quadrats' residuals vary—this is not unexpected as the number of data points per quadrat is small—but the minimum average risk still leads to a reasonable penalization.

In a CV procedure the model needs to be fitted K times so computational cost and CV risk estimate stability are to be balanced. In the spatial setting an additional problem arises due to border correction, i.e. we use $W_k \ominus b(o, r_{\mathrm{bor}})$ instead of W_k in equation (12) with some $r_{\mathrm{bor}} > 0$. Since each subwindow needs to be reduced by the same border correction range as the original window to obtain truly independent subsets of data, the loss of data for estimating $\hat{R}_{\gamma,k}$s limits the number of splits that can be done on W, as some data are simply lost in the process. For example, in a square window, a border correction range that is 5% of the window's side together with a 3×3 partitioning will effectively lead to a 50% loss of data when estimating the risk (Fig. 2). Furthermore, the varying abundances of the types need weighting to account for importances

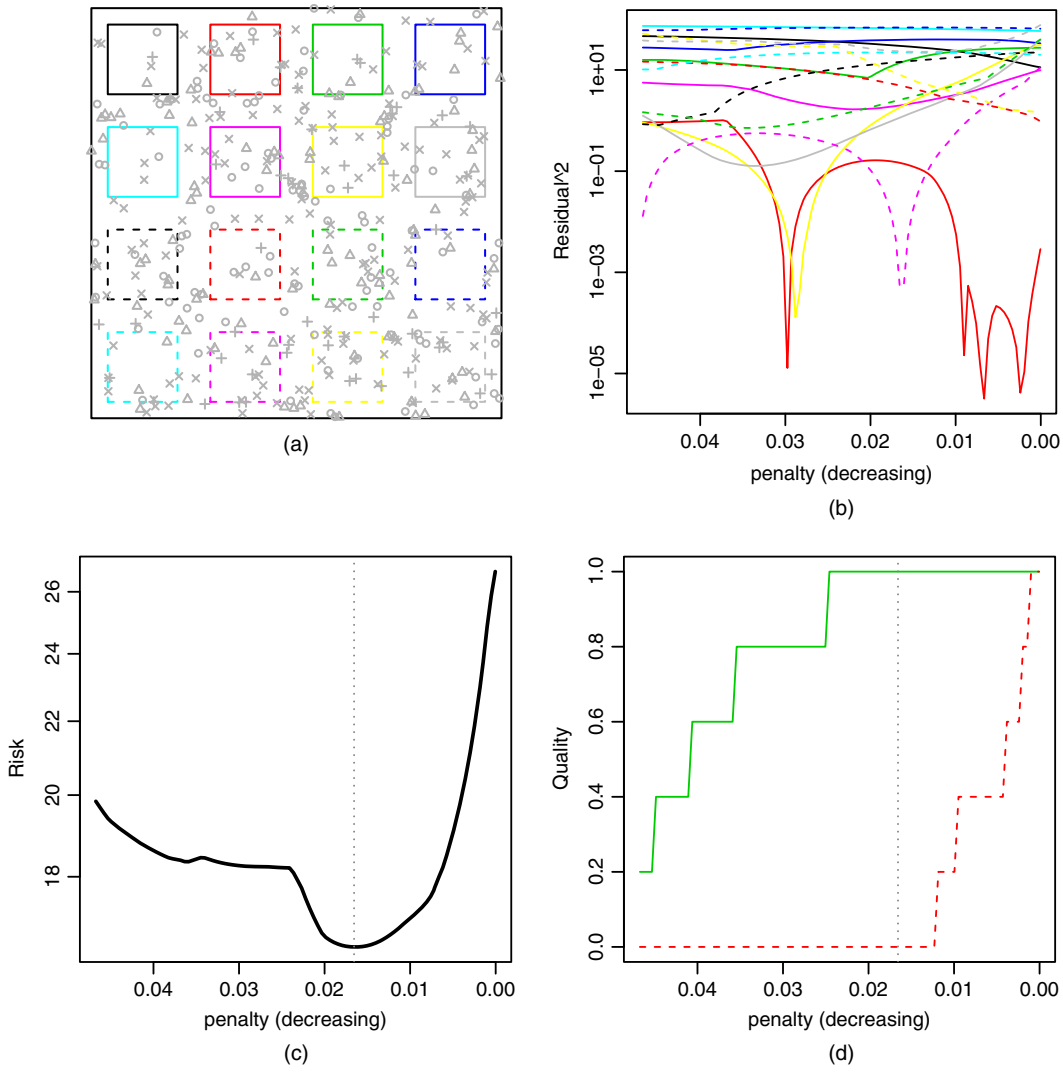

Fig. 1. Example of the CV penalty selection: (a) a pattern and the CV partitioning; (b) the squared inverse residuals $\hat{R}^2_{\gamma,k}$ per quadrat $k = 1,\ldots,16$; (c) the estimated prediction risk $\hat{R}_{\mathrm{CV}}(\gamma)$; (d) true (———) and false (– – –) positive rates per γ (⋮, penalty level that would be chosen by the method)

of different types, and heterogeneity of data should be considered as well. It is therefore very difficult to give general guidelines for partitioning the window that would work in all scenarios.

For comparison with the CV penalty selection, we also include a rule-of-thumb penalty selection based on the Akaike information criterion (AIC). We computed the AIC for the group lasso as described by Breheny and Huang (2009). In our experiments, penalties with the lowest AIC, say γ_{AIC}, consistently led to too dense solutions. This suggests that either the pseudolikelihood approximation or the spatial dependence, or both, leads to underestimation of the effective number of parameters. As a plug-in rule of thumb, initial trials indicate that a penalty around $(\gamma_{\mathrm{AIC}} + \gamma_{\max})/2$ leads to reasonable penalization; we shall report these results with the label AIC0.5.

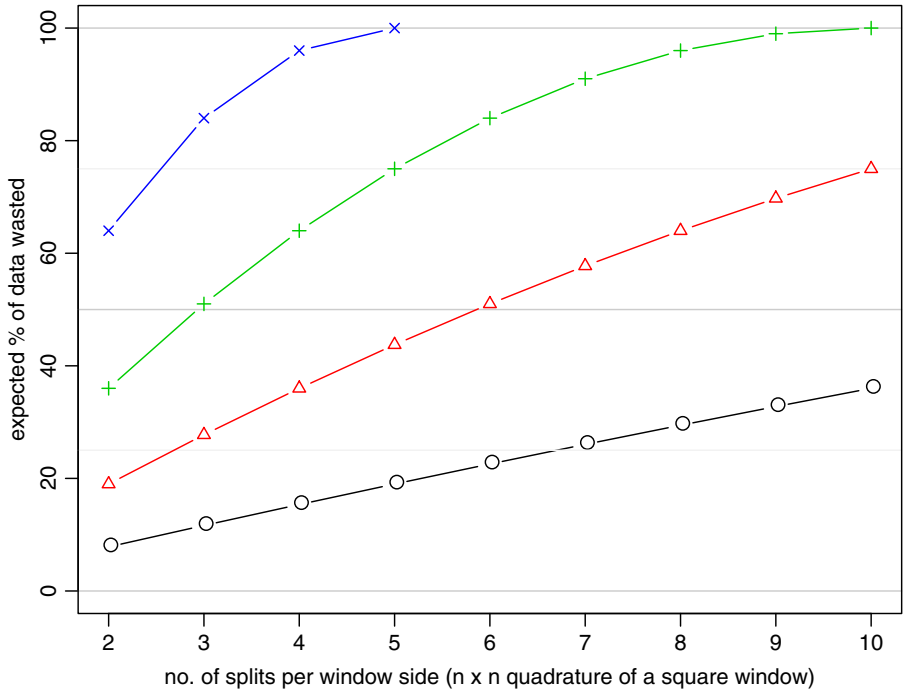

Fig. 2. Loss of data when doing CV splits with border correction with range R in a square window: ×, 10%; +, 5%; △, 2.5%; ○, 1%

We also constructed a Bayesian algorithm for computing approximate posteriors $\Pr(e_g = 1|\mathbf{x})$. We call this tailored algorithm the variational Bayes spike-and-slab (VBSS) algorithm. It uses a quadratic approximation to the logistic function (Jaakkola and Jordan, 2000), selects variables via the spike-and-slab priors (Mitchell and Beauchamp, 1988) and uses the mean field variational approximation to infer the maximum *a posteriori* estimates (Ormerod and Wand, 2010). In experiment 1 we also checked a non-grouped version of the VBSS algorithm and compared it with the Markov chain MC based spike-and-slab generalized additive model (SSGAM) (Scheipl, 2011) that has a slightly different implementation of the spike-and-slab priors. All of these alternatives solve the same optimization problem as the group lasso (equation (11)), but with different choices of penalization.

3. Alternative approach for multitype interaction detection: non-parametric Monte Carlo testing

The current state of the art methodology for analysing high dimensional multivariate point pattern data is based on non-parametric summary statistics and MC testing. This approach splits the analysis into a set of bivariate tests for no coassociations (second-order interactions). As a reference to our proposed method we apply an MC testing scheme in the BCI rainforest-like, large inhomogeneous simulation experiment of Section 4.5 and the actual BCI rainforest data analysis in Section 5. We give here a short description of MC testing but the interested reader should seek more detailed texts on the topic; see for example Illian *et al.* (2008), section 7.5, Baddeley *et al.* (2014), Velázquez *et al.* (2016) and Brown *et al.* (2016).

MC-based testing procedures in general, and tests of spatial type-to-type interactions in particular, are built on three components:

(a) a null hypothesis that describes a counterfactual pattern without the interactions of interest;
(b) a statistical summary that measures the interaction of interest;
(c) a statistical test based on the summary for measuring data's departure from the distribution of the summary when the null hypothesis holds.

In the context of potentially inhomogeneous patterns, a popular null model, especially in ecology, is some form of the inhomogeneous Poisson process (IPP) (Wiegand *et al*., 2012). A common asymmetric version can be simulated by keeping type i locations fixed while sampling type j locations from the IPP independently from all other data. The idea is to nullify short-range interactions but to keep longer-range, environment-related associations via the spatially varying intensity. The intensity surface that is needed for the IPP simulations of $\mathbf{x}_j$ is estimated with a fixed smoothness that reflects understanding of the division between 'short' and 'long' ranges. We implement the IPP simulations following Wiegand *et al*. (2012): for each type $i = 1, \ldots, p$, an intensity field $\eta_i(u)$, $u \in W = [0, 1000]\ \mathrm{m} \times [0, 500]\ \mathrm{m}$ is kernel estimated on a $2\ \mathrm{m} \times 2\ \mathrm{m}$ grid using the border-corrected Epanechnikov kernel with bandwidth fixed to 30 m (this is of course not optimal for all types; we simply emulate Wiegand *et al*. (2012)). Then n_i points are distributed on W with density relative to $\eta_i(u)$, i.e. we use a conditional IPP as is commonly done. We simulate 999 patterns this way for each type $i = 1, \ldots, p$.

For the statistical summary of pattern interaction, a popular choice is the Ripley cross-K_{ij} function. It describes the number of type j neighbours for an average point of type i, over different neighbourhood ranges. So, assuming isotropy, for each pair of bivariate patterns the $K_{ij}(r)$ function estimate is a sampled curve over spatial scales $r > 0$. In our analyses, we estimate the cross-type $K_{ij}(r)$ curves on a range grid $r = r_1, \ldots, r_{\max}$ with translation edge correction. For each species $i = 1, \ldots, p$ and $j = 1, \ldots, p$, we estimate the curve $K_{ij}^0 = \{K_{ij}^0(r_1), \ldots, K_{ij}^0(r_{\max})\}$ from the bivariate data pattern $\mathbf{x}_i \cup \mathbf{x}_j$. When $i = j$ we estimate Ripley's univariate $K_{ii} = K$. Subsequently, we estimate the curves K_{ij}^b from the bivariate synthetic patterns $\mathbf{x}_i \cup \mathbf{x}_j^b$ where $\mathbf{x}_j^b$ is a simulation of the null model for j as described above, and $b = 1, \ldots, 999$. With the set of curves $(K_{ij}^0, \ldots, K_{ij}^{999})$ for all p^2 combinations of i and j, we do the variance stabilizing $\sqrt{\{K(r)/\pi\}}$ transform to increase statistical power.

The third component of the MC testing framework is then needed for a proper combined test for the K_{ij}-curves, i.e. to determine whether K_{ij}^0 is different from the null model curves. A family of such tests is called deviation tests or envelope tests (Myllymäki *et al*., 2017; Baddeley *et al*., 2014), and several options are available. We shall use the Studentized deviation test, which measures the L^2-distance of the Studentized curves (scaled with respect to the null model), and the rank envelope test, which is a multi-dimensional analogue of a rank test (see Myllymäki *et al*. (2017)). Each test leads to a p-value per test, say p_{ij}. We then report the values $e_{ij} := \mathbf{1}(p_{ij} < 0.05)$ as indicators of interaction.

This concludes our description of our methodological framework and explains our automated approach to the selection of 'active' interactions that are important to explain the observed spatial pattern.

4. Simulation trials

To check the model fitting procedure before data analysis we illustrate its characteristics via multiple simulation trials. We shall focus on the estimation of the cross-interaction terms, which we can write as a square matrix

$$M = [M_{ij}]_1^p, \qquad M_{ij} = \begin{cases} e_{g(ij)} & i \neq j, \\ e_{g(i)} & i = j. \end{cases}$$

We call M the interaction matrix. Note that the coefficients that are related to each interaction function $\beta_{ij}^{\mathrm{T}}\mathbf{g}_{ij}$ are in groups, meaning that either the whole step function is estimated to be $\hat{\beta}_{ij} = 0$, or one or more of the components β_{ijk} is estimated to be non-zero. This is an example of the group lasso (Yuan and Lin, 2006).

The key quality metrics that we shall look at to assess group level performance are the true positive rate TP and the false positive rate FP, stratified to intratype and intertype interactions when relevant.

We simulated patterns in either a $W = [0, 10]^2$ or $W = [0, 1000]\ \mathrm{m} \times [0, 500]\ \mathrm{m}$ window with different settings for interactions. Experiments 1, 2 and 4 study the method under the scenario of a correctly specified model, i.e. we simulate and estimate the multirange multivariate model. In experiments 3 and 5 we simulate Cox models to see how general and flexible the step function (shrinkage) approach can be for interaction discovery. Simulations of our model were carried out by using the birth-and-death algorithm for Fig. 12 in Appendix C, but for the trials we shall use a Metropolis–Hastings algorithm with fixed point counts to keep intensities at desired levels (for more detail we refer to Illian *et al*. (2008), pages 147–154).

4.1. Experiment 1: interactions in a small pattern

We first simulated a small ($p = 4$) example to familiarize the reader with our methodology and to check that the method works as intended. We simulated 100 realizations of the multirange multivariate saturation model with per-type point counts (100, 100, 50, 150) in a $[0, 10]^2$ window. The saturation levels were all set to $c = 1$, producing low levels of interaction. Intratype ranges were set to $\mathbf{r}_i = (0.1, 0.2, 0.3)$, and intertype ranges were set to $\mathbf{r}_l = (0.1, 0.4)$. Types 1 and 2 were set to exhibit internally a mixture of short-range repulsion and medium-range clustering ($\beta_{ii} = (-1, 1, 0),\ i = 1, 2$), type 3 had some medium-range clustering ($\beta_{33} = (0, 1, 0)$) and type 4 had no internal correlation. A positive intertype correlation ($\beta_{ij} = (0.6, 0.3)$) was set between types 1 and 2 and types 3 and 4. The true range vectors were used for fitting. A strongly penalizing hyperprior $\Pr(e_g = 1) \sim \mathrm{beta}(0.1, 10)$ was chosen for the Bayesian methods. A 4×4 partitioning of W was used for CV to keep the data loss around 50%.

Fig. 3 depicts the average detection rates per coefficient β_{ijk}. Apart from Pearson residuals, the rates of the grouped methods are very similar, indicating that the estimated effects do not depend on the algorithm chosen (the group lasso or spike and slab). The non-zero structure in types 1 and 2 is detected well by the method. The ungrouped SSGAM and VBSS algorithms capture well the medium-range clustering for types 1 and 2 but not the short-range regularity. This indicates that the medium-range clustering is the main signal at the group level. The type 3 subpatterns had interaction only at medium range, and it seems to be difficult to uncover by any of the methods that were used. We posit that this difficulty is due to a smaller point count and mixed interaction with type 4. The grouping helps to discover the intertype interactions between the 1–2 and 3–4 pairs, which are evident from the lower detection rates of the ungrouped methods.

A summary of the true and false positive rates at group level for all methods is given in Table 1. The ungrouped outputs are considered non-zero per group if any of the group members were estimated non-zero. We used a threshold of 0.5 for the Bayesian posterior probabilities for classification. The raw and Pearson residuals clearly are too prone to false positive results in this example to be useful.

We also studied how varying the range vectors and dummy intensities affect the outcome. The maximum ranges in the simulations were $R = 0.3$ for intratype interactions and $R = 0.4$ for

Fig. 3. Illustration of the variable selection results for experiment 1 (three different fitting algorithms (the group lasso, SSGAM and VBSS algorithms, as described in Section 2.4) (symbols denote coefficientwise averages over 100 simulations; jittered grey crosses in the background denote the per-simulation estimates; if a large symbol is closed, the true coefficient was non-zero; the grouping is denoted by vertical lines; the corresponding type(s) with numbers (e.g. intratype interaction for type $i = 3$ by '3' and intertype interaction for $i = 1$ and $j = 3$ by '1-3')); penalty selection for the group lasso based on CV with (a) inverse and (b) Pearson residuals; (c) AIC-based rule-of-thumb penalty AIC0.5; (d) VBSS algorithm with beta(0.1, 10) penalty prior, 'grouped'; (e) SSGAM; (f) VBSS algorithm with beta(0.1, 10) penalty prior, 'ungrouped'

Table 1. Experiment 1 true positive and false positive rates, mean and standard deviations (in parentheses) over 100 simulations

Parameter	*Result for the following methods:*						
	AIC0.5	*CV inverse*	*CV Pearson*	*CV raw*	*SSGAM*	*VBSS grouped*	*VBSS ungrouped*
TP	0.77 (0.18)	0.78 (0.26)	0.95 (0.15)	0.99 (0.08)	0.78 (0.17)	0.83 (0.17)	0.80 (0.16)
FP	0.03 (0.07)	0.10 (0.20)	0.51 (0.32)	0.94 (0.16)	0.02 (0.06)	0.21 (0.19)	0.12 (0.15)

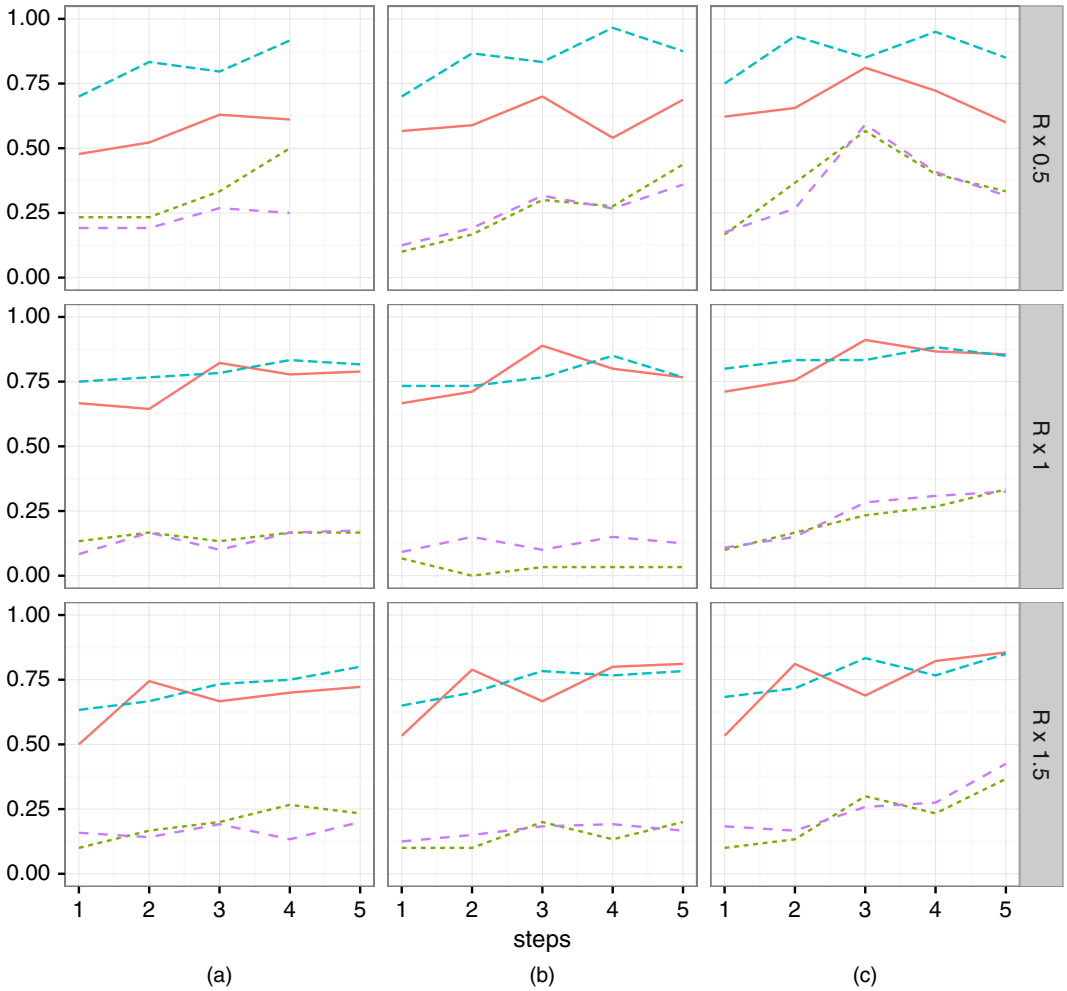

Fig. 4. Experiment 1 (group lasso with AIC0.5 penalty), how varying the range vectors $\mathbf{r} = (r_1, \ldots, r_{\text{steps}} = R_{\max})$ and dummy intensity factor affects the quality of group lasso interaction detection quality (in the simulations the intratype interaction maximum range was $R = 0.3$ and the intertype $R = 0.4$, and the number of steps was 3 (intratype) and 2 (intertype)) (———, TP intratype; ------, FP intratype; — — —, TP intertype; — — —, FP intertype): (a) $\rho = 2$; (b) $\rho = 4$; (c) $\rho = 8$

Table 2. Experiment 2 interaction detection rates, stratified by intratype and intertype interactions and point count per type†

Method	*Results for $n_i = 50$*			*Results for $n_i = 100$*			*Results for $n_i = 200$*		
	Intra TP	*Inter TP*	*Inter FP*	*Intra TP*	*Inter TP*	*Inter FP*	*Intra TP*	*Inter TP*	*Inter FP*
Raw	0.51 (0.20)	0.97 (0.04)	0.04 (0.06)	0.65 (0.21)	0.99 (0.02)	0.03 (0.05)	0.89 (0.10)	1.0 (0.00)	0.20 (0.23)
Inverse	0.41 (0.22)	0.96 (0.05)	0.04 (0.07)	0.63 (0.37)	0.95 (0.11)	0.15 (0.27)	0.17 (0.33)	0.76 (0.19)	0.06 (0.21)
Pearson	0.90 (0.13)	0.99 (0.02)	0.55 (0.36)	0.97 (0.05)	1.0 (0.01)	0.78 (0.35)	0.99 (0.04)	1.0 (0.00)	0.90 (0.27)
AIC0.5	0.00 (0.01)	0.61 (0.11)	0.00 (0.00)	0.00 (0.00)	0.64 (0.10)	0.00 (0.00)	0.00 (0.00)	0.57 (0.07)	0.00 (0.00)
VBSS (1,1)	0.85 (0.12)	0.95 (0.04)	0.50 (0.13)	0.97 (0.07)	0.99 (0.02)	0.91 (0.20)	0.91 (0.09)	1.0 (0.00)	0.86 (0.25)
VBSS (0.1,1)	0.37 (0.16)	0.58 (0.27)	0.23 (0.12)	0.97 (0.08)	0.85 (0.19)	0.87 (0.22)	0.90 (0.11)	1.0 (0.02)	0.46 (0.48)
VBSS (0.1,10)	0.09 (0.11)	0.14 (0.22)	0.07 (0.07)	0.98 (0.07)	0.62 (0.19)	0.77 (0.24)	0.84 (0.20)	0.97 (0.07)	0.34 (0.45)

†Mean over 50 simulations, with the standard deviation given in parentheses; intra FPs are 0 by design; lasso penalty selected with 25-fold CV.

intertype interactions. To simulate misspecification of the ranges, we altered these by multiplying by a factor 0.5 or 1.5 and created range vectors for fitting with one, two, three, four or five equidistant steps. Then we fitted the model by using the group lasso with the AIC0.5 rule for the penalty with dummy intensity factor ρ either 2, 4 or 8.

Two main features emerged from these studies (Fig. 4). First, choosing ranges that are too short reduces the quality of detection quite markedly. Second, the number of grid steps in the interaction functions can be misspecified without large variation in the results, giving a type of numerical robustness. Many small steps in the interactions functions are not recommended, because of the small number of pairs in the data that hit every annulus. For dummy intensity factor 2 the joint point configuration did not have enough point pairs to fill every bin in the range grid of five steps, even after resampling the dummies (10 repeated attempts). As we also saw in the results for type $i = 3$ above, using many steps may result in lower true positive rates at group level because the group lasso penalizes the whole group equally over all its members. The false positive rates are higher for dummy intensity factor $\rho = 8$ than for $\rho = 4$, which is a side effect of the AIC0.5 rule of thumb. A check of the AIC curves showed that the minimum was often achieved with lower penalty when $\rho = 8$, probably because there are more observations in the logistic regression design matrix, which is an unsurprising indication that the standard model selection tools are not to be trusted when using pseudolikelihood. Since we are focusing on the grouped analysis in this study we continue with the group lasso and VBSS algorithm.

4.2. Experiment 2: finding interactions in blocks

Next, we increased the type count to $p = 10$ and added interactions in two blocks with in-block pairwise interactions but no interaction between the blocks. The main task was to discover the two blocks with minimal amount of false positive results in their cross-section.

The ranges for simulating from the multirange saturation model were all set to a two-step vector $\mathbf{r} = (0.25, 0.50)$, and saturations fixed to $c = 1$. The first block of five types had short- and medium-range clustering, $\beta_{ii} = (1, 0.5)$ for $i = 1, \ldots, 5$; the second block of five types had short-range repulsion with mild medium-range clustering, $\beta_{ii} = (-1, 0.5)$ for $i = 6, \ldots, 10$. The intertype interaction for each pair in both blocks was positive correlation with $\beta_{ij} = (0.5, 0.25)$, with no correlation between blocks. We simulated three intensity scenarios, having point count per type n_i either 50, 100 or 200, so that the total point counts per simulated pattern were either 500, 1000 or 2000. The window was again $[0, 10]^2$.

Fitting was done with the misspecified range vectors $\mathbf{r} = (0.15, 0.3)$, to increase the challenge (see experiment 1). CV was conducted with a 5×5 partitioning to keep the expected data loss around 50%. The VBSS algorithm was fitted with three hyperpriors $\pi_l \sim \text{beta}(\cdot, \cdot)$ ranging from flat (1,1) to medium (0.1, 1) and strong (0.1, 10) preference for no interactions. The choices correspond to increasing the penalization in the group lasso and facilitate comparisons between the methods.

The rates of interaction detection are shown in Table 2. The VBSS algorithm produces many false positive results, even with the strong prior that should penalize towards sparsity. The raw CV method works well, as does the inverse CV method, but the Pearson CV has a high false positive rate. AIC0.5 overpenalizes, thereby missing all of the intratype interactions, indicating that the rule of thumb is not generally useful.

Fig. 5 depicts the interaction matrix estimate for a single realization and the mean interaction over the simulations when $n_i = 100$, as given by the VBSS algorithm with the strong prior and inverse CV method. We plot the main diagonal of the interaction matrix from the south-west corner to the north-east corner, situating the $(1, 1)$ pixel at the south-west corner, as per Fig. 1 in Flügge *et al.* (2014). Note that white indicates that the detection rate was 1.

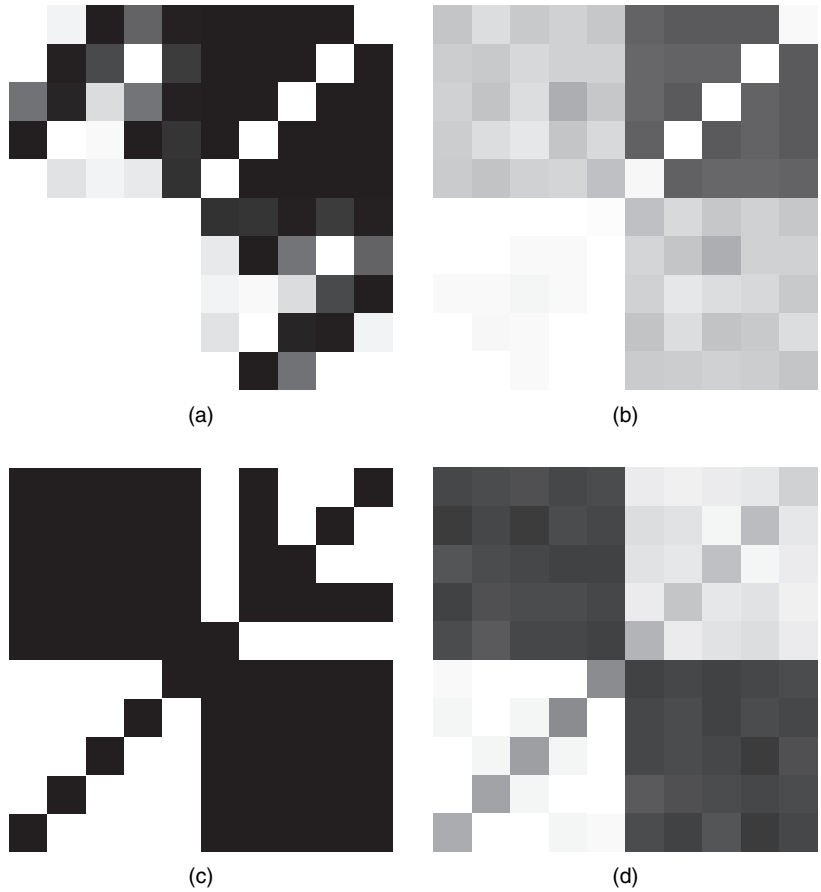

Fig. 5. Results for experiment 2 (shown are an example interaction matrix estimate and the mean interaction matrix, for both VBSS and inverse CV methods; the point count per type was 100; each pixel (i, j) corresponds to $M_{i,j}$; black means that the interaction detection rate was 0, and white means that it was 1): (a) first simulation, VBSS(0.1, 10); (b) average, VBSS(0.1, 10); (c) first simulation, inverse; (d) average, inverse

We note that the first block is clear in VBSS matrices but the second block, with intratype regularity mixing with the intertype clustering, is not so clear and many false positive results have been detected. The group lasso is not so efficient in detecting the intratype interactions but the intertype interactions are much better detected, with only a few false positive results on average.

Further tailoring of ranges will improve quality, as per experiment 1. Adjusting the priors helps the VBSS algorithm but provides little practical gain over the automatic penalty rules for the group lasso which work reasonably well on all cases. As the priors lead to varying results, analysis with the VBSS algorithm needs to include a sensitivity analysis, similar to lasso penalization selection. The major issue is that instead of having one tuning parameter there are several, and cross-validating the space becomes infeasible as (at least with our implementation) the VBSS algorithm is not much faster for a single prior choice than the group lasso algorithm is for the whole penalty path. The usual Bayesian benefits such as posterior distributions and variances are not worth pursuing either because of the likelihood approximation. For these practical reasons we shall continue with only the group lasso to perform computationally feasible variable selection.

Table 3. Experiment 3, log-Gaussian Cox processes, intertype interactions, average and standard deviation over 50 simulations

Method	*Results for corr 0.6*				*Results for corr 0.8*			
	$n_i = 50$		$n_i = 100$		$n_i = 50$		$n_i = 100$	
	TP	*FP*	*TP*	*FP*	*TP*	*FP*	*TP*	*FP*
Raw	0.60 (0.12)	0.16 (0.16)	0.60 (0.16)	0.16 (0.20)	0.62 (0.13)	0.08 (0.14)	0.65 (0.12)	0.07 (0.15)
Inverse	0.60 (0.10)	0.14 (0.12)	0.64 (0.10)	0.17 (0.11)	0.67 (0.08)	0.11 (0.13)	0.73 (0.13)	0.17 (0.15)
Pearson	0.75 (0.11)	0.41 (0.24)	0.71 (0.09)	0.31 (0.18)	0.77 (0.08)	0.32 (0.19)	0.80 (0.07)	0.31 (0.20)
AIC0.5	0.20 (0.06)	0.00 (0.00)	0.22 (0.06)	0.00 (0.00)	0.23 (0.07)	0.00 (0.00)	0.27 (0.08)	0.00 (0.00)

4.3. Experiment 3: detecting blocks of interacting log-Gaussian Cox processes

Next we check whether interactions can be detected in correlated log-Gaussian Cox process data. We do not expect large power here because of the local Poisson distribution of the patterns, but it is important to check how flexible the detection is under model misspecification as the rainforest data potentially exhibit a variety of spatial mechanisms.

We simulated homogeneous, stationary multivariate log-Gaussian Cox processes of $p = 24$ types, structured into three correlated blocks of eight types each. Inside a block the eight latent Gaussian fields are correlated linear model of co-regionalization fields (Gelfand *et al.*, 2002), for which we set the cross-field correlation levels to either 0.6 or 0.8. All fields were driven by a Matérn covariance function with smoothness $\nu = 10$, marginal variance $\sigma^2 = 2$ and correlation range 3 (so that $\text{corr}(r > 3) < 0.1$). For illustration, four subpatterns from two different blocks are plotted in Figs 6(a)–6(d) overlaid on their generating fields. We ran the experiment with constant $n_i = 50$ and $n_i = 100$ points per type. The window was $[0, 10]^2$.

Estimation range vectors for intratype and intertype interactions were set to $\mathbf{r} = (0.25, 0.5)$ after examining one realization's cross-pair correlation function. The CVs were carried out using a 3×3 partitioning to keep the expected data loss at 50%.

Table 3 lists the detection rates. The block structure is detected to some extent, with somewhat elevated false positive rates. The raw residual CV performs the best, Pearson CV resulting in high false positive rates, and inverse CV landing overall somewhere in between. The AIC0.5 rule penalizes too much to detect more than approximately 25% of the block structure. As expected, an increase in interfield correlation improves the detection, as does doubling the point count. Figs 6(d)–6(g) depict the estimated interaction matrices for the raw CV when point counts were $n_i = 100$. Even though the intertype interactions are not always detected correctly, the intratype clusterings are detected quite well.

We noted that improvements are easily achieved by adjusting the range vectors that were used in the fitting, but eyeballing a single realization's pair correlation function is obviously not optimal for a repeated experiment. A practical solution is to choose the ranges per realization on the basis of a proper exploration of the data, in which case we expect quite a good performance considering how misspecified the model is for the generating mechanism. During our development of the methods, further improvements in this experiment were also achieved by increasing the level of clustering in the patterns, either by increasing the variance or by decreasing the range. For example, setting $\sigma^2 = 3$ (the original level was $\sigma^2 = 2$) reduced false positive rates and

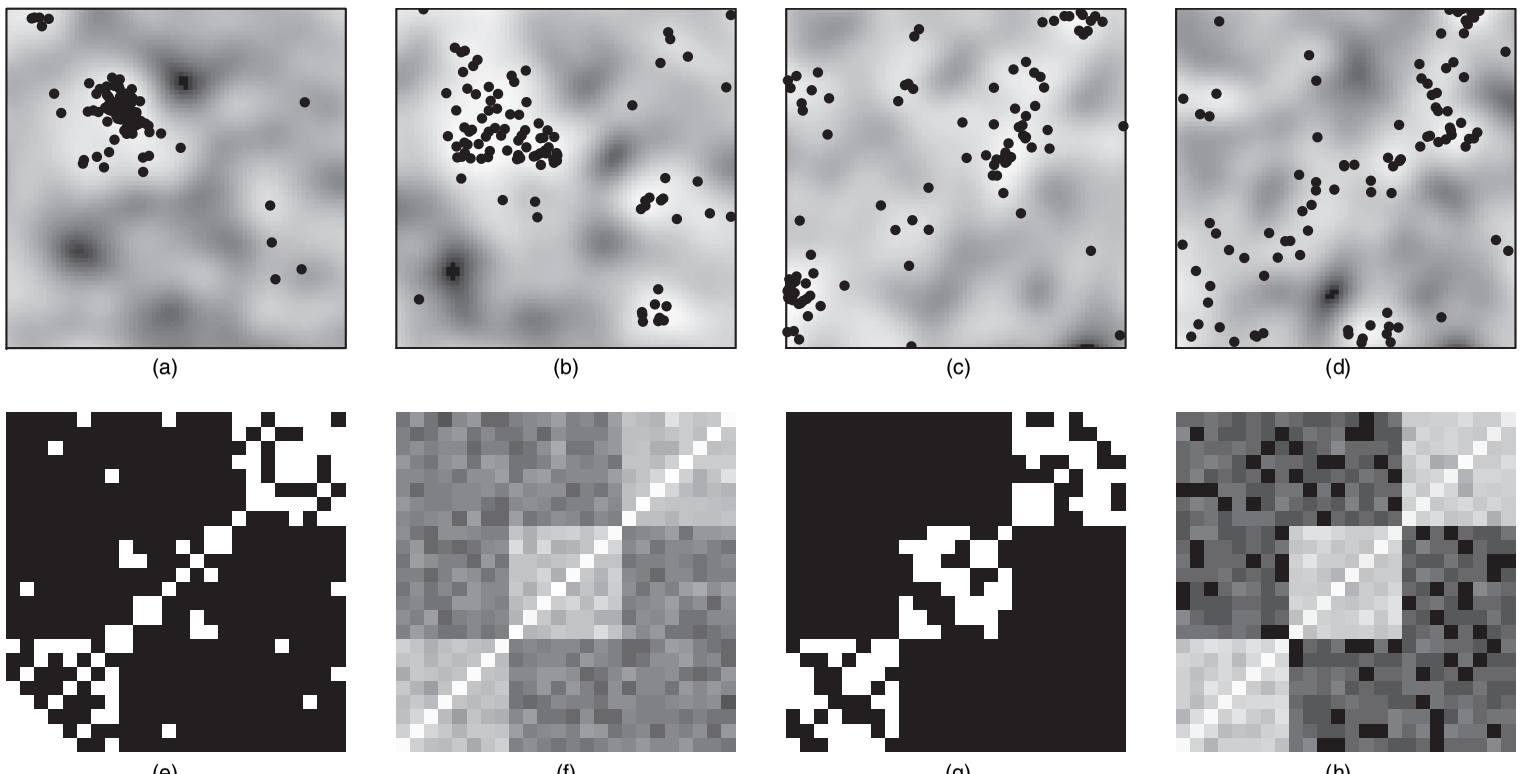

Fig. 6. Experiment 3, block-correlated log-Gaussian Cox processes ($n = 100$): (a)–(d) examples of the point patterns in a simulation, two patterns from each of the two intracorrelated blocks, overlaid on their generating log-Gaussian fields; (e)–(h) example and the average interaction matrix for two levels of intrablock correlation, estimated by the raw residual CV method (black means that the interaction detection rate was 0, and white means that it was 1; note that the (1,1) pixel is at the south-west corner of the image); (a) example type 1, block 1, corr = 0.8, (b) example type 2, block 1, corr = 0.8; (c) example type 9, block 2, corr = 0.8; (d) example type 10, block 2, corr = 0.8; (e) example interaction matrix, corr = 0.6, raw CV; (f) mean interaction matrix, corr 0.6, raw CV; (g) example interaction matrix, corr = 0.8, raw CV; (h) mean interaction matrix, corr = 0.8, raw CV

Table 4. Experiment 4, detection rates with and without habitat effects and mostly no interaction†

	Results for $p=10$				*Results for $p=20$*			
	Raw	*Inverse*	*Pearson*	*AIC0.5*	*Raw*	*Inverse*	*Pearson*	*AIC0.5*
Homogeneous								
FP intra	0.44 (0.42)	0.07 (0.12)	0.08 (0.15)	0.15 (0.12)	0.19 (0.20)	0.02 (0.04)	0.04 (0.07)	0.08 (0.06)
FP inter	0.47 (0.39)	0.06 (0.09)	0.12 (0.19)	0.18 (0.07)	0.22 (0.19)	0.03 (0.04)	0.07 (0.10)	0.10 (0.03)
Homogeneous + 1 interaction								
FP intra	0.13 (0.27)	0.01 (0.03)	0.14 (0.17)	0.00 (0.00)	0.08 (0.20)	0.04 (0.10)	0.06 (0.10)	0.00 (0.00)
FP inter	0.07 (0.13)	0.00 (0.01)	0.13 (0.16)	0.00 (0.00)	0.06 (0.17)	0.02 (0.06)	0.06 (0.07)	0.00 (0.00)
TP	1.0 (0.00)	1.0 (0.00)	1.0 (0.00)	1.0 (0.00)	1.0 (0.00)	1.0 (0.00)	1.0 (0.00)	1.0 (0.00)
Inhomogeneous								
FP intra	0.25 (0.36)	0.11 (0.13)	0.19 (0.34)	0.14 (0.14)	0.13 (0.09)	0.00 (0.02)	0.02 (0.03)	0.10 (0.08)
FP inter	0.22 (0.34)	0.08 (0.10)	0.18 (0.31)	0.13 (0.10)	0.10 (0.07)	0.01 (0.01)	0.01 (0.01)	0.06 (0.03)
Inhomogeneous + 1 interaction								
FP intra	0.08 (0.19)	0.02 (0.06)	0.13 (0.22)	0.00 (0.00)	0.12 (0.24)	0.00 (0.00)	0.04 (0.05)	0.00 (0.00)
FP inter	0.05 (0.08)	0.01 (0.02)	0.13 (0.20)	0.00 (0.00)	0.12 (0.21)	0.00 (0.01)	0.04 (0.06)	0.00 (0.00)
TP	1.0 (0.00)	1.0 (0.00)	1.0 (0.00)	1.0 (0.00)	1.0 (0.00)	1.0 (0.00)	1.0 (0.00)	1.0 (0.00)

†When the type count p is doubled, the interpair count is quadrupled ($\times$ 4.22).

increased true positive rates up to 10% across Table 3. The results presented were chosen as a generic illustration, and, by eye at least, the simulation parameters produced clustering that was similar to that observed in the rainforest data.

4.4. Experiment 4: rare interactions

To focus more specifically on the false positive rates of the candidate methods, we simulated patterns where, at most, there is only one pairwise intertype interaction that is non-zero. We considered four scenarios:

(a) homogeneous multivariate Poisson patterns,
(b) homogeneous multivariate Poisson patterns with an extra intracorrelated pattern,
(c) inhomogeneous multivariate Poisson patterns and
(d) inhomogeneous multivariate Poisson patterns with an extra intracorrelated pattern.

For this and the following experiment 5 we used real covariate maps from the BCI data set. 13 different soil variables (such as soil acidity and magnesium concentration), together with elevation and elevation gradient maps, are available on a 20 m $\times$ 20 m grid covering the observation window $W = [0, 1000]$ m $\times$ $[0, 500]$ m. For this experiment (and the data analysis later on) we decomposed the 15 covariate maps by using a singular value decomposition of the pointwise measurement matrix, and kept the six component maps corresponding to the six largest singular values, capturing approximately 70% of the features. These six pointwise independent PCA maps were then used as covariates.

To generate trends for the inhomogeneous multivariate Poisson processes, we combined the principal component analysis (PCA) covariates linearly with random coefficients taking values -1, 0 or 1, with probabilities 0.25, 0.5 and 0.25 respectively. For each simulation of the multivariate process, two trend maps were generated and exponentiated to work as intensity surfaces. Each of the two intensity surfaces was assigned to half of the p subprocesses. This way each

simulated multivariate pattern has two blocks of types, $p/2$ each, with very strong intrablock correlation that is fully explained by the covariates. For each simulation the point counts per type were set to range from 50 to 300, increasing log-linearly so that lower counts were more common. We repeated the experiment for $p = 10, 20$.

The extra intracorrelated patterns were generated with the multitype multiscale model with homogeneous intensity of 100 points and exhibited very short-range repulsion and medium-range clustering ($\mathbf{r} = (1, 20)$; $\beta = (-10, 1)$; $c = (3, 3)$). For fitting we set all ranges to $\mathbf{r} = (7, 15)$ m.

The results for the four scenarios are given in Table 4. With the Poisson data, the methods all produce some false positive results. Inverse CV is best, with around 6% false positive rates. With the exception of Pearson CV, including the extra interacting type sharpens the results by reducing the false positive rates. Every method discovers the extra interaction perfectly. Including covariate dependence, and producing correlation, does not increase the false positive rates overall. Again, the inclusion of a single interacting type reduces the false positive rates while itself being clearly detected. Doubling the type count roughly quadruples the parameter count, but it does not seem to affect the quality as false positive rates go down, albeit not always by a factor of 4.

In the estimation procedure we did not penalize the covariate coefficients, because we found that penalizing them by using the group lasso leads to underprediction of the trend due to the shrinkage effects of the lasso, consequently leading to underpenalization by the CV. When the covariate effects are to be considered more closely we suggest a bias correction step to the lasso or the use of a less strongly penalizing added term such as smoothly clipped absolute deviation or the minimax concave penalty (Breheny and Huang, 2009).

4.5. Experiment 5: independent patterns in a rainforest landscape

As a final synthetic example we simulated $p = 64 = 4 \times 16$ independent, inhomogeneous patterns in a $W = [0, 1000]\ \mathrm{m} \times [0, 500]\ \mathrm{m}$ window with first-order interaction depending on the covariates in the BCI data. We selected half (2×16) of the patterns to come from a Thomas cluster process (Illian *et al.* (2008), section 6.3.2) with two different dispersal ranges and points-per-cluster rates, to arrive at an appropriate degree of heterogeneity. We set the first block's model to generate patterns with a few large clusters ('Thomas 1'), and the second block's model to generate patterns with many small clusters ('Thomas 2'). The second half of the patterns was generated by using the multitype multirange model with either repulsion followed by clustering ('Geyer 1') or just repulsion ('Geyer 2'). The ranges of the multirange models were inversely dependent on the target point count: this way small amounts of points could spread out more realistically, and if the target point count was high then maximal packing density would not be violated (i.e. only a finite amount of 'discs' with certain radius fit inside W).

To simulate habitat effects we connected the patterns to the maps of the four covariates *Mn*, *P*, *pH* and *grad* which have been found to be relevant covariates for the rainforest population (Schreeg *et al.*, 2010). The covariate values were first standardized. Then for each simulated species we sampled uniformly $t \in \{0, \ldots, 4\}$ of them and summed them pointwise with weights $(t, \ldots, 1)/(t + 1)$ to produce a surface. The surface was then used as the unscaled intensity field for Thomas process generator points, and as first-order field for the multistep multivariate Geyer models. The covariate choices and the coefficients were kept fixed over the replicates of the multivariate simulations.

In a single realization of the process, each of the four models were simulated 16 times, with target intensities within each 16-type block ranging from 50 to 1000 points, median 225. The realized point counts ranged from 34 to 1024 because of edge effects while simulating the Thomas

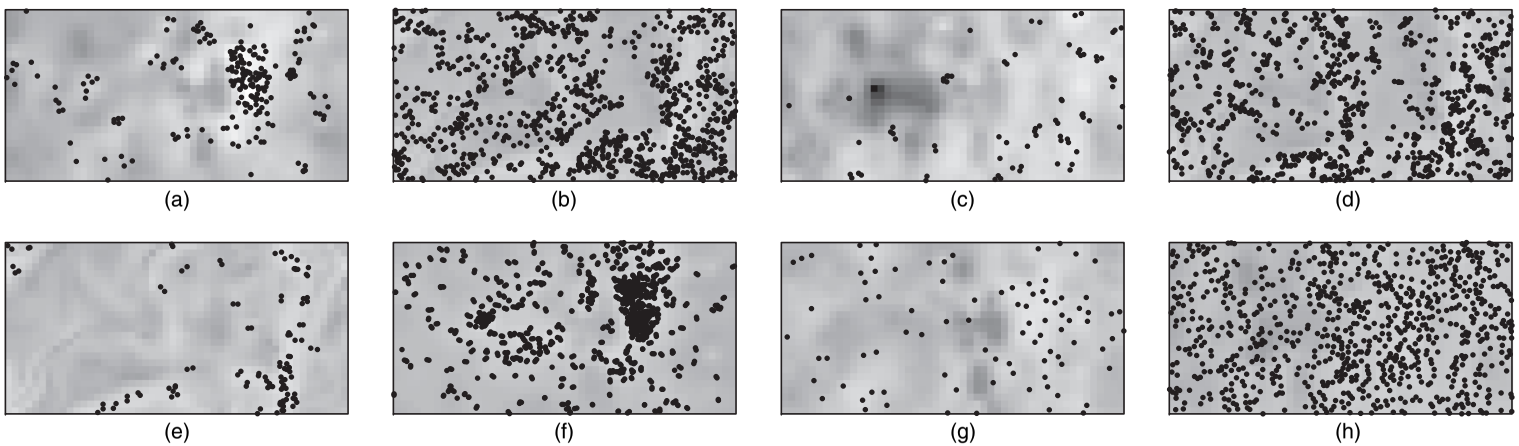

Fig. 7. Experiment 5, examples of the subpatterns with non-zero habitat association in a single multivariate realization, together with the habitat effect field (window $[0, 1000]$ m $\times$ $[0, 500]$ m): (a) type 7, model Thomas 1, $n = 209$; (b) type 16, model Thomas 1, $n = 1006$; (c) type 21, model Thomas 2, $n = 77$; (d) type 32, model Thomas 2, $n = 859$; (e) type 37, model Geyer 1, $n = 110$; (f) type 48, model Geyer 1, $n = 1000$; (g) type 52, model Geyer 2, $n = 90$; (h) type 63, model Geyer 2, $n = 820$

Table 5. Experiment 5, synthetic forest of independent species: detection rates for the new approach, together with results for two MC tests, Studentized deviance and rank envelope, with two testing ranges†

	CV raw	*CV inverse*	*CV Pearson*	*AIC0.5*
Gibbs model				
TP	0.76 (0.34)	0.91 (0.06)	0.96 (0.08)	0.35 (0.06)
FP	0.01 (0.01)	0.05 (0.10)	0.20 (0.12)	0.00 (0.00)
	Studentized deviance (0.5–15 m)	*Studentized deviance (1–50 m)*	*Rank envelopes (0.5–15 m)*	*Rank envelopes (1–50 sm)*
MC test				
TP	0.74 (0.05)	0.89 (0.03)	0.80 (0.05)	0.96 (0.01)
FP	0.02 (0.00)	0.00 (0.00)	0.03 (0.00)	0.02 (0.00)

†By design, all 64 intratype interactions were non-zero, and all 2016 intertype interactions were zero; average and standard deviation over 10 simulations.

process (points outside W were dropped). The point count in a single multivariate realization was around 20 800.

Fig. 7 depicts eight subpatterns in a typical realization, each overlaid on its inhomogeneity-generating habitat field. Thomas 1 patterns have about half the number of clusters of the Thomas 2 patterns. Note that it is difficult to see from the plots the very short-scale features operating at ranges 1–10 m as the area is so large.

For fitting we used ranges $\mathbf{r} = (10, 20)$ and 7×4 partitioning to keep the estimated data loss at 50%. Table 5 shows the detection rates for the proposed methods together with the MC-based Studentized deviance and rank envelope test, both based on the cross-K functions estimated over ranges 0.5–15 m and 1–50 m (corresponding to short- and medium–long-range testing scenarios) with 999 simulations of the IPP null model, as described in Section 3. Apart from the Pearson CV, and the AIC0.5 rule of thumb which penalizes too much, the detection rates are good. Around or below 5% of the false intertype interactions are detected whereas 74–96% of the intratype interactions are detected. The model-based and the MC-based estimates are very close to each other.

Somewhat surprisingly, the Thomas patterns, for which the model is misspecified, are not the more difficult of the two families to discover. Fig. 8 shows the averages and examples of interaction matrix estimates for the MC tests and raw and inverse CV methods. We placed the (1,1) interaction at the south-west corner of each image. The matrices are ordered by model (Thomas 1, Thomas 2, Geyer 1 and Geyer 2) and increasing point count from the bottom left to the top right within the model blocks. For intratype interactions (on the diagonal), the rarer species's interaction within the Geyer 2 model seems to be more difficult to discover with the CV methods. This could be due to the fixed $\mathbf{r}$ that was used for estimation when the processes had varying ranges. The MC null hypothesis design, where the types corresponding to columns in the interaction matrix were kept fixed and the types corresponding to rows were randomized, leads to non-symmetric estimates. This is an important feature of the MC method and, as we return to in the data analysis below, can lead to interpretational issues. The Geyer 1 block, where there is short-scale repulsion, and medium-scale attraction, is the major source of false

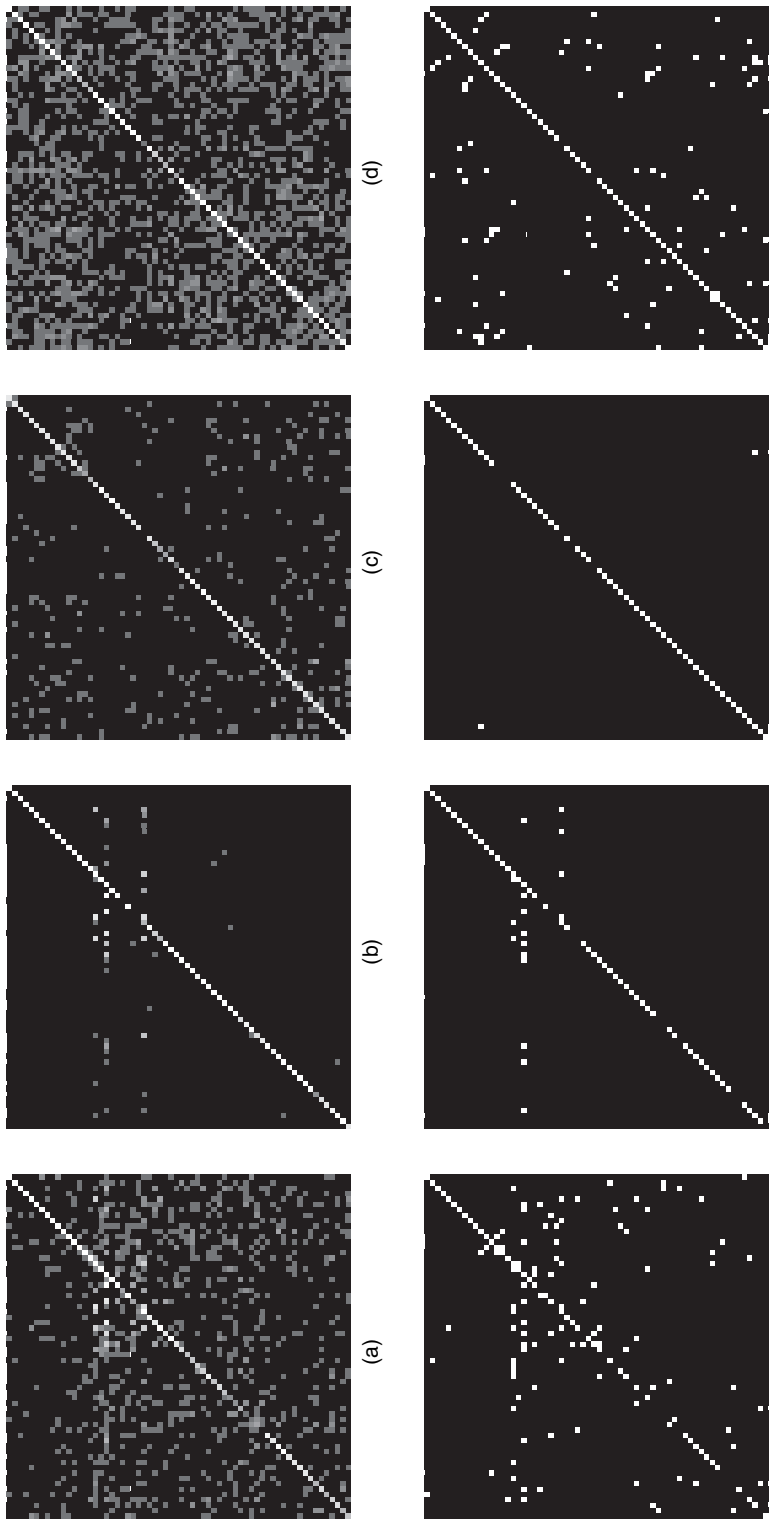

Fig. 8. Experiment 5, (a)–(d) averages and (e)–(h) examples of estimated interaction matrices (the results from two MC tests with two different range scale coverages (see the text) and the raw and inverse CV methods are shown; a perfect estimate would be a diagonal matrix; Thomas models are in the lower left-hand section and Geyer models in the top right-hand section of each matrix): (a) MC Studentized deviance test, 0.5–15 m; (b) MC Studentized deviance test, 1–50 m; (c) raw CV; (d) inverse CV; (e) MC Studentized deviance test, 0.5–15 m; (f) MC Studentized deviance test, 1–50 m; (g) raw CV; (h) inverse CV

positive results for the MC tests, and types with higher abundance show generally higher false positive rates. This is potentially important, because such a mixture of repulsion and attraction is likely to be common in many plant communities.

4.6. Summary of the simulation trials

The results from the simulation studies are very encouraging despite the sometimes large dimensionality of the underlying problems, especially demonstrated with experiment 5. In experiment 1 we established feasibility of our model, and the agreement between penalization methods validated the approach. Experiment 2 demonstrated the model's ability to discover block community structures in a moderately complicated 10-type pattern. Experiment 3 showed that the model can detect significant interactions in not just realizations of itself, but also in realizations of Cox processes. Experiment 4 showed that the false positive rate does not need to be high for us to be able to discover rare events. Finally in experiment 5 we saw that the method handles rainforest-like patterns with covariate effects, varying point counts and various interaction mechanisms and models well.

The experiments suggest that CV with Pearson residuals is not reliable in practice. The raw residuals work well in some cases (experiments 2 and 3) but fail completely in others. The inverse residuals produced most consistently good detection. When the raw and inverse residuals both worked well, the variances of the inverse residuals were higher, so computing both is advisable.

As noted in the experiments, we did not specifically tailor the range parameters for each fitting task *per se*. Adjusting the model's range scales according to specific analysis is highly recommended as they are expected to improve the results in practice.

5. Data example: rainforest interactions

The BCI rainforest data set consists of multivariate point patterns corresponding to censuses of rainforest plants (shrubs and trees) living in a $W = [0, 1000]\ \mathrm{m} \times [0, 500]\ \mathrm{m}$ area of BCI, Panama. Censuses have been taken regularly since 1981 (Hubbell *et al.*, 2005; Condit, 1998; Condit *et al.*, 1999). In each census, woody plants (shrubs and trees) with diameter at breast height over 1 cm were catalogued, noting their diameter at breast height, location, species, condition and some other details that are not relevant to our studies in this paper.

The total number of species is about 300, with slight variation over the years due to immigration and extinctions in W, which is physically an open area allowing for migration. For the example data analysis we chose the 2005 census and selected some specific species. Recent studies (Kanagaraj *et al.*, 2011; Yang *et al.*, 2016) suggest that spatial features of the plants vary with life stage. Normally the distribution of young plants is more clustered because of seed dispersal and adult plants' distribution more regular because of competition-based self-thinning, and that resource requirements also change with maturation. We therefore selected a subset of species for which an estimate of reproducible size, which is a surrogate for juvenile or adult life stage, was available (unpublished data by R. Foster, available as supplements for Flügge *et al.* (2014)). We included only adult plants and excluded species with fewer than 50 adults in the region, leading to a multivariate pattern of $p = 83$ species. Point counts vary from 50 to 8784, with a median of 118 and a total of 31 650.

5.1. Interaction detection using the methods introduced

We carried out the interaction estimation with a multirange multivariate saturation model using group lasso penalization as in experiment 5. Habitat effects were accounted for by including the

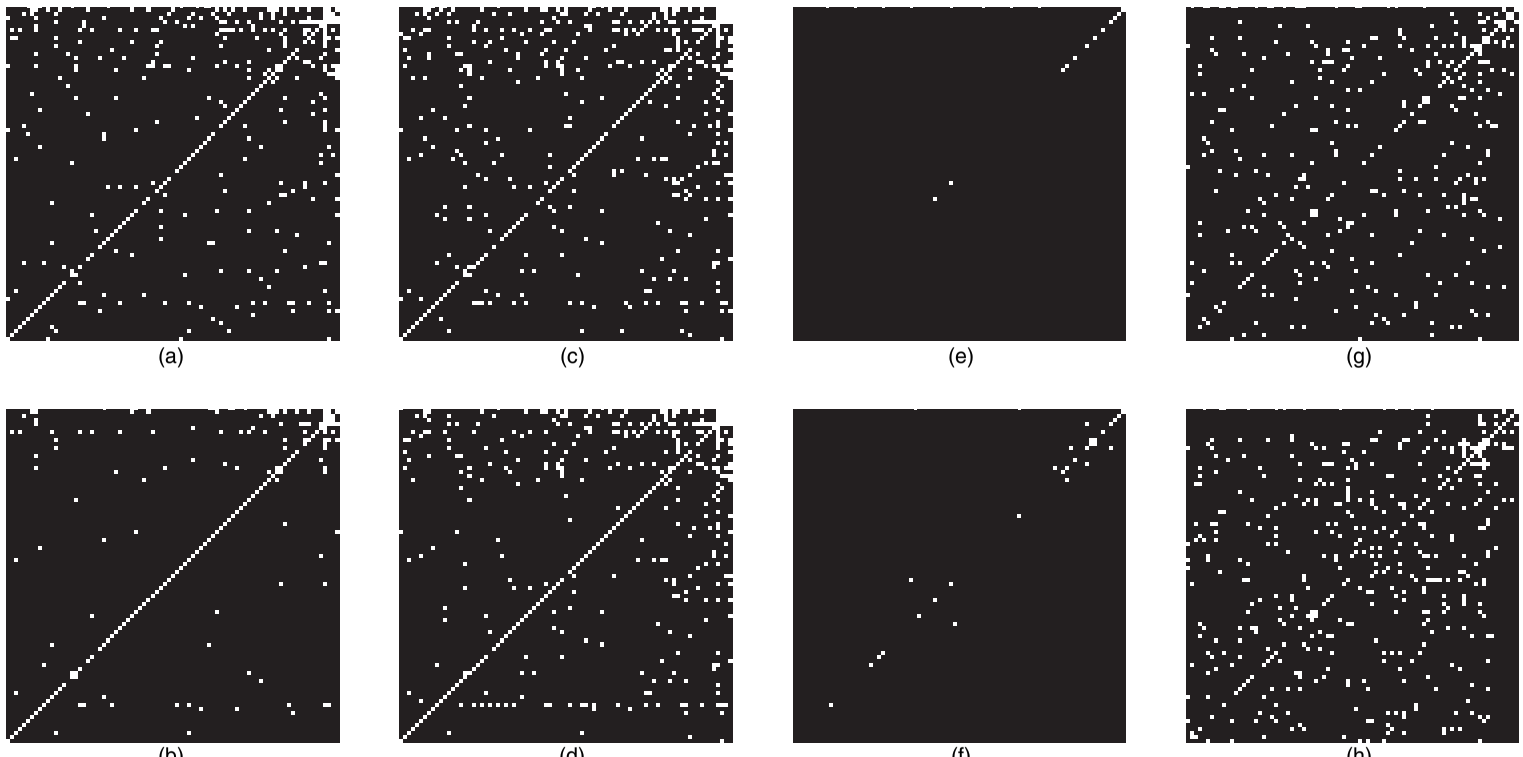

Fig. 9. Different estimates of the interaction matrix of the adult plant BCI 2005 census: (a)–(d) non-parametric MC method with two different MC tests and two different spatial range vectors (the species kept fixed in the test is on the x-axis, and the randomized species is on the y-axis; (e)–(h) group lasso with raw residual CV and inverse residual CV methods, and two range vectors (species are arranged by point count, increasing from the bottom left to the top right); (a) Studentized deviation test, ranges 0.5–15 m; (b) ranges 0.5–30 m; (c) rank envelope test, ranges 0.5–15 m; (d) ranges 0.5–30 m; (e) raw residual CV, ranges 7 m and 15 m; (f) ranges 7 m, 15 m and 30 m; (g) inverse residual CV ranges 7 m and 15 m; (h) ranges 7 m, 15 m and 30 m

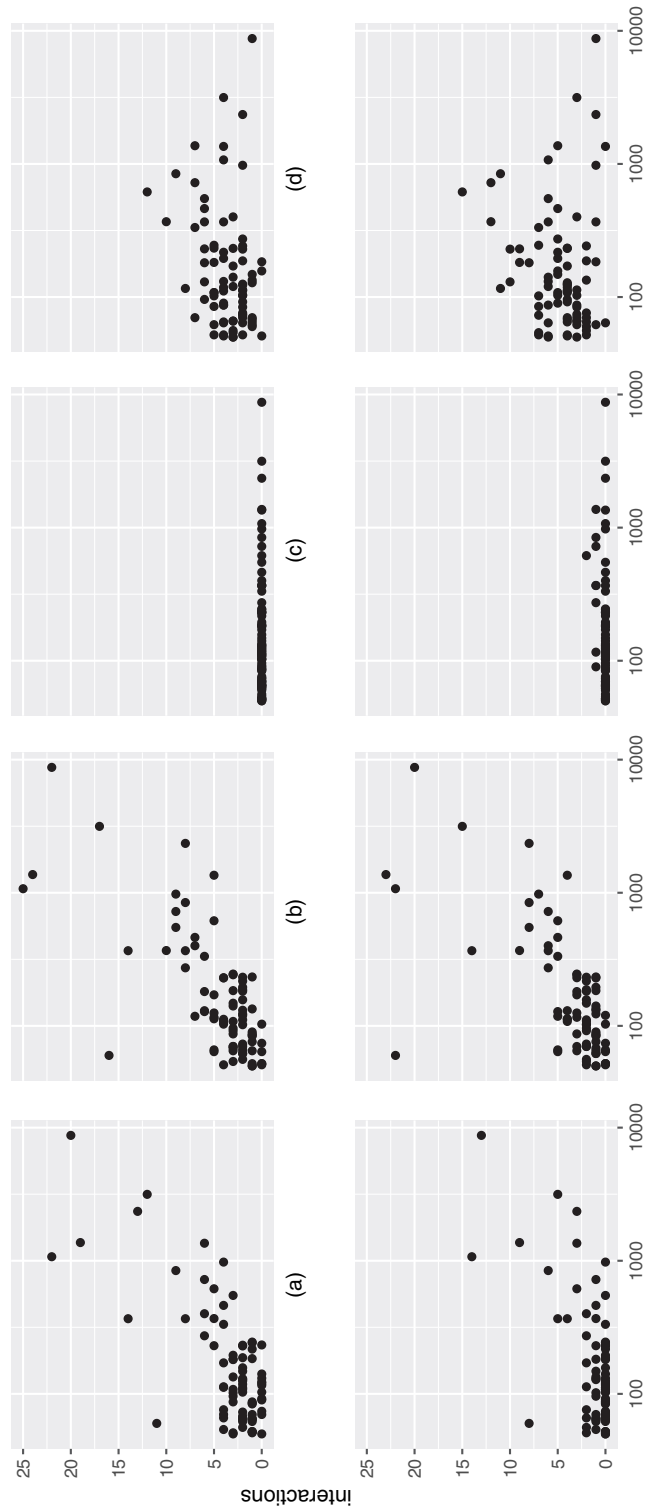

Fig. 10. BCI 2005 adults analysis, number of detected interspecies (*ij*) interactions for each species *i*, as a function of abundance: ranges up to (a)–(d) 15 m and (e)–(h) up to 30 m; (a), (e) MC, Studentized deviance; (b), (f) MC, rank envelopes; (c), (g) CV raw; (d), (h) CV inverse

Table 6. Intratype and intertype interaction rates estimated with the proposed method using raw and inverse CV, and MC tests using the Studentized deviance and rank envelopes†

	Results for the following methods:			
	Studentized deviance	*Rank envelopes*	*CV raw*	*CV inverse*
Up to 15 m				
Intratype	0.89	0.88	0.14	0.55
Intertype	0.05	0.06	0.00	0.04
Up to 30 m				
Intratype	0.95	0.96	0.33	0.60
Intertype	0.02	0.05	0.00	0.06

†Results for range scales up to 15 m and 30 m are given separately. $p = 83$ species; 3403 distinct pairs.

six PCA maps as covariates as described in experiment 4 (using PCA maps avoids problems with collinearity of covariates). We set the range vector equal across all 3486 intratype and intertype interactions and fitted the model twice, with range vector $\mathbf{r} = (7, 15)$ m and with range vector $\mathbf{r} = (7, 15, 30)$ m, the ranges being concordant with previous results of neighbourhood-dependent growth models within the BCI forest (Table 4, Uriarte *et al.* (2004)). We used the dummy intensity factor 4, with a minimum of 500 dummies per type to avoid singularities. The saturation parameters are set according to Section 2.3. To correct for edge effects we implemented a 30-m buffer zone, and, to keep expected data loss below 50%, the CV partitioning was fixed to 6×3. For comparison, we also implemented the MC tests as described in Section 3.

5.2. Results

The resulting interaction matrices for the two MC tests are depicted in Figs 9(a)–9(d). The group lasso estimates with raw and inverse residual CV penalty selection are depicted in Figs 9(e)–9(h). Figs 9(a), 9(c), 9(e) and 9(g) involve two ranges up to 15 m, whereas Figs 9(b), 9(d), 9(f) and 9(h) involve three ranges up to 30 m. The MC test interaction matrices are not symmetric by design. For example, according to the rank envelope test at ranges 0.5–30 m, species $i = 10$ interacts significantly with one species when the other species in a pair is randomized in the test, but it interacts significantly with 22 species when itself is the one randomized in the test. Although biological interactions (competition) can be asymmetric, with one species being a superior competitor to a second species, the spatial correlations that emerge from these interactions should be symmetric. The non-symmetric spatial interaction matrices that are produced by the MC method are therefore problematic when it comes to interpretation. In comparison, the inverse CV model-based method estimates the number of interactions for species $i = 10$ to be 5, of which four are among the 22 species that are indicated by the MC test. Another difference is that the MC test outcome suggests that more abundant species interact more. The model-based results do not indicate such a trend (Fig. 10); so, when the numbers of intertype interactions are of similar order, the distributions are different. Finally, nearly all species are deemed internally structured by the MC test, but around or less than half by the model-based approach (Table 6).

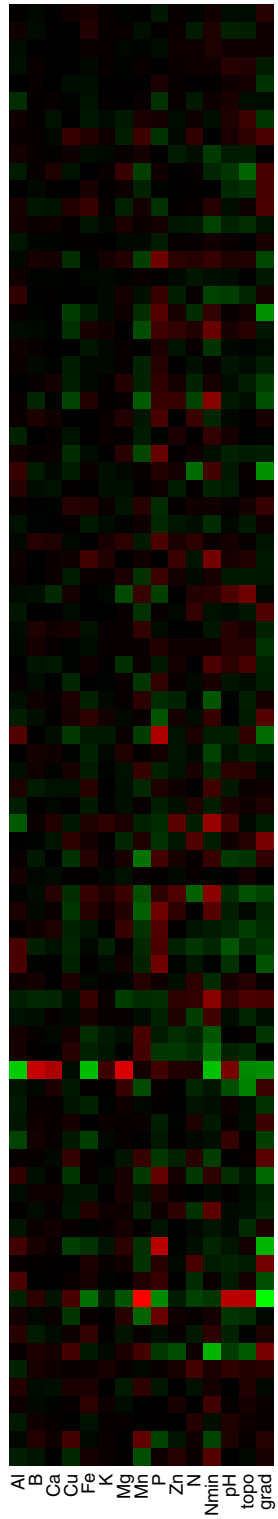

Fig. 11. Estimated covariate effects with the raw CV method when the range vectors are $\mathbf{r} = (15, 30)$ m: positive connections are green: negative are red; brightness reflects magnitude (black≡0); range of values $[-0.9, 1.03]$; the estimates do not vary much between the methods or the range vectors used because the covariates were not penalized

We have designed a complete analysis pipeline that can be used to answer some specific ecological questions, and in particular to uncover important processes affecting the forest. For example, we could now proceed by analysing the covariate effects (Fig. 11) to see how important environment (soil and elevation) effects are in structuring the forest. Then, to consider structure caused by second-order effects, we could further study the interaction matrices to discover sub-communities and interacting groups as was done by Flügge *et al*. (2014) and Morueta-Holme *et al*. (2016) based on different MC approaches.

6. Discussion

We have developed a flexible model-based method for detecting small-scale interactions in multivariate spatial point patterns, demonstrated its potential on several synthetic examples and applied it to a large rainforest data set. This is in contrast with using log-Gaussian Cox processes for multivariate modelling, as in for example Waagepetersen *et al*. (2016), where small-scale interactions are more difficult to capture. In so doing we have greatly extended the potential applicability of model-based inference for multitype point patterns, allowing both for the analysis of more interactions and for more complex multiscale behaviour.

In the Thomas cluster process, and in Cox processes in general, the two-stage generation of 'offspring' given 'mothers' is a natural model for plant regeneration in natural plant populations. However, as the offspring are assumed mutually independent, the model is suitable for only a particular type of pattern, such as just germinated individuals. The individual-to-individual competition and survival over a plant's lifetime before adulthood is better captured with explicit pointwise interaction models, such as Gibbs models. Alternative models that emulate natural thinning in addition to natural clustering can be constructed (Stoyan, 1979; Andersen and Hahn, 2016; Lavancier and Møller, 2016), but fitting such models relies heavily on non-parametric methods for which simultaneous multivariate analysis is currently not well understood. Our approach takes a more data analytical approach by including all effects simultaneously and addressing the high dimensionality with penalized inference.

The model can be tailored for specific applications with relatively low effort since apart from the point-to-point indicator functions the analysis pipeline is fixed. For example, adding Strauss components might detect clustering more efficiently as the saturation level of the Geyer components caps the clustering tendency. Such a model would not work as a generative model and simulations would not be stable (singular clusters would form; Gates and Westcott (1986)), but it could be effective as a statistical interaction discovery tool.

The pipeline is also computationally practical. Group lasso algorithms are very fast and work well with computational underlying calculations for sparse matrices, which the approximation design matrices for Gibbs models are by construction. On a regular laptop it took us days to compute all MC estimates to compare with our model, but it took only hours to estimate the full model. Of course, one can parallelize the MC estimation since the type-to-type interaction tests are performed independently.

An advantage of our model-based approach is the explicit treatment of covariates that may affect the intensity but not the interactions directly. This means that it is possible to infer which covariates are important for the distributions of the different species. In contrast the MC approach that predominates in the ecological literature relies on inhomogeneous Poisson processes to capture the first-order (environmental) processes, and the user needs to determine or estimate the scale over which these processes are affecting the intensities. The current trend has been to use one scale for all species, but this is unlikely to be optimal. However, we note that such covariates are not always available, and the MC method might be prefer-

able when covariates have not been measured but are thought to be important. The within-type spatial structure that is commonly found in the data is also broken down when conducting the null model test in the MC approach, and it is not clear what effect this might have on results. Nonetheless the user must also make important decisions in our model-based approach, such as the range of scales over which to carry out the analysis, and as we discuss below the model validation and penalization methods also are an area of future exploration.

The computational burden of model fitting was increased tenfold because we determined the best penalty level by using CV. To simplify the penalty selection, we tried including a few extra Poisson 'noise' subpatterns in the data to select the penalization so that the noise stayed independent. The result of this approach was inconclusive and remains an area of future investigation. We also tried using CV with the constructed design matrix of the logistic regression likelihood, as is common for lasso applications but, much like with the AIC, too many false positive results resulted for this approach. Standard model selection approaches do not do well because the likelihood is an approximation, rather than the correct form of the likelihood. Correcting the score and Hessian of the pseudolikelihood while doing penalization would be useful in this regard, and some work has already been done in the unpenalized case by Coeurjolly and Rubak (2013). Using a corrected Hessian would also enable us to do better inference on the covariates, as (approximate) confidence intervals could then be constructed.

Replacing the group lasso with some other penalization could have several benefits. Some penalizations shrink non-zero coefficients less, which is ideal both for the predictions that are needed in the CV step and for downstream analysis of the estimated effects. For example, we suspect that the shrinkage issue prevented the method from working properly when covariates were penalized and consequently the inhomogeneous trends became too flat. A further limitation of the group lasso is that all group members were penalized equally; this in turn led to low group detection with many range steps and a low amount of data per range annuli. A more refined group penalization method such as sparse group lasso or minimax concave penalty is more sensitive to individual group members being non-zero (Breheny and Huang, 2009) and would enable more detailed interaction functions with more steps. Adding steps does, however, require the use of more dummy points so that each annulus registers something and numerical problems are avoided. We could also replace the step functions with overlapping components, such as radial basis functions, and gain not only numerical stability but also smoother interaction function estimates. Further refinement to the CV optimization penalty grid is also needed in actual applications. In our high p examples the 100-step grid was often too coarse to include or exclude individual groups.

The model is constructed via small point-to-point interactions, and any large-scale unobserved variability is not equally well captured. It might be possible to add spatial random effects, such as Gaussian processes, to the first-order interaction and still to use the same pipeline. This would bridge the interface between Gibbs models and log-Gaussian Cox processes, which is an exciting area of future investigation. This paper therefore stands as a further step to understand general heterogeneous and high dimensional point process observations.

We would also like to point out that the modular construction of the point-to-point interaction functions provides a potentially useful connection to dynamic modelling of ecological communities defined through generations of dispersal-related birth and competition-based death events (Law and Dieckmann, 2000; Law *et al.*, 2003). If the transition probabilities are modelled log-linearly, the stationary distribution (if it exists) of the corresponding birth-and-death process

following the detailed balance condition would be a Gibbs process of the kind that we have discussed here in the more general, potentially inhomogeneous and highly multivariate setting.

7. Software and data

All computations were implemented in R (version 3.3.1). The random fields for the log-Gaussian Cox models were simulated by using the R package `RandomFields` (version 3.1.3), and uniform simulation and some utilities were used from the R package `spatstat` (version 1.46.1). For the group lasso we adapted the local co-ordinate descent algorithm in the R package `grpreg` (version 3.0-2) with the inclusion of the offset terms. The VBSS algorithm was implemented by hand. For the SSGAM we used the R package `spikeSlabGAM` (version 1.1-11), with hyperprior beta distribution settings as for the VBSS algorithm and otherwise using default parameters. An R package implementing the method pipeline is available from the first author.

The BCI data are available for research purposes from `http://ctfs.si.edu/webatlas/datasets/bci/`.

Acknowledgements

The BCI forest dynamics research project was founded by S. P. Hubbell and R. B. Foster and is now managed by R. Condit, S. Lao and R. Perez under the Center for Tropical Forest Science and the Smithsonian Tropical Research Institute in Panama. Numerous organizations have provided funding, principally the US National Science Foundation, and hundreds of field workers have contributed. We thank the Engineering and Physical Sciences Research Council for support via grant EP/N007336/1 and EP/L001519/1, and S. C. Olhede also acknowledges support from the 7th European Community framework programme (a European Research Council Fellowship via grant CoG 2015- 682172NETS).

We thank the two reviewers for their very helpful comments.

Appendix A: Details about the likelihood approximation

Let X be the point process generating the pattern $\mathbf{x}$ and let D be a marked dummy point process in the window W which is independent of X and has known intensity functions ρ_i, $i = 1, \ldots, p$. Baddeley *et al.* (2014) proposed to solve the likelihood optimization problem by using the estimating function

$$s_W(X, D; \theta) := \sum_{u \in X \cap W} \frac{\rho(u)\mathbf{v}(u; X \setminus u)}{\lambda_\theta(u; X \setminus u) + \rho(u)} - \sum_{u \in D \cap W} \frac{\mathbf{v}(u; X)\lambda_\theta(u; X)}{\lambda_\theta(u; X) + \rho(u)}, \tag{13}$$

where $\rho(u) = \rho_i(x)$ for $u = (x, i)$. With the help of the Campbell formula and the Georgii–Nguyen–Zessin theorem (formulas 1.5.10 and 6.6.2 in Illian *et al.* (2008)), it can be shown that s_W is an unbiased estimation function and that finding the root of s_W gives an unbiased estimate of the maximum of likelihood (3).

The s_W as a function of θ is the derivative of

$$\log\{\tilde{f}_\theta(X \cap W)\} = \sum_{u \in X \cap W} \log\left\{\frac{\lambda_\theta(u; X \setminus u)}{\lambda_\theta(u; X \setminus u) + \rho(u)}\right\} + \sum_{u \in D \cap W} \log\left\{\frac{\rho(u)}{\lambda_\theta(u; X) + \rho(u)}\right\}, \tag{14}$$

which formally is the likelihood of a logistic regression with variables $\tau(u) = \mathbf{1}(u \in X)$ for $u \in X \cup D$ and

$$P\{\tau(u) = 1\} = \frac{\lambda_\theta(u; X \setminus u)}{\lambda_\theta(u; X \setminus u) + \rho(u)} = \frac{\exp[\theta^{\mathrm{T}}\mathbf{v}(u; X) + \log\{\rho(u)^{-1}\}]}{1 + \exp[\theta^{\mathrm{T}}\mathbf{v}(u; X) + \log\{\rho(u)^{-1}\}]}.$$

In practice the method works as follows: we sample a set of dummy points $\psi = \{(x, t)\} = \cup \psi_i$ with subpatterns ψ_i having a known distribution with constant intensity ρ_i in W, for each type $i = 1, \ldots, p$. We then

calculate the vectors $\mathbf{b}(u) := \mathbf{v}(u; \mathbf{x})$ for each $u \in \mathbf{x} \cup \psi$. Let $N = \#(\mathbf{x} \cup \psi)$. The log-likelihood of the logistic regression can then be written in a compact vector form corresponding to

$$\log\{\tilde{f}_\theta(\mathbf{x})\} = \sum_{u=(x,i)\in(\mathbf{x}\cup\psi)\cap W} (t(u)\{\theta^{\mathrm{T}}\mathbf{b}(u) + o(u)\} - \log[1 + \exp\{\theta^{\mathrm{T}}\mathbf{b}(u) + o(u)\}]) \tag{15}$$
$$= \mathbf{t}^{\mathrm{T}}(B\theta + \mathbf{o}) - \mathbf{1}_N^{\mathrm{T}} \log\{\mathbf{1}_N + \exp(B\theta + \mathbf{o})\},$$

where $t(u) = \mathbf{1}(u \in \mathbf{x})$ indicate the data points, B is a row matrix of $\mathbf{b}(u)$s for each $u \in \mathbf{x} \cup \psi$ and $o(u) = -\log\{\rho_i(u)\}$ are offset terms.

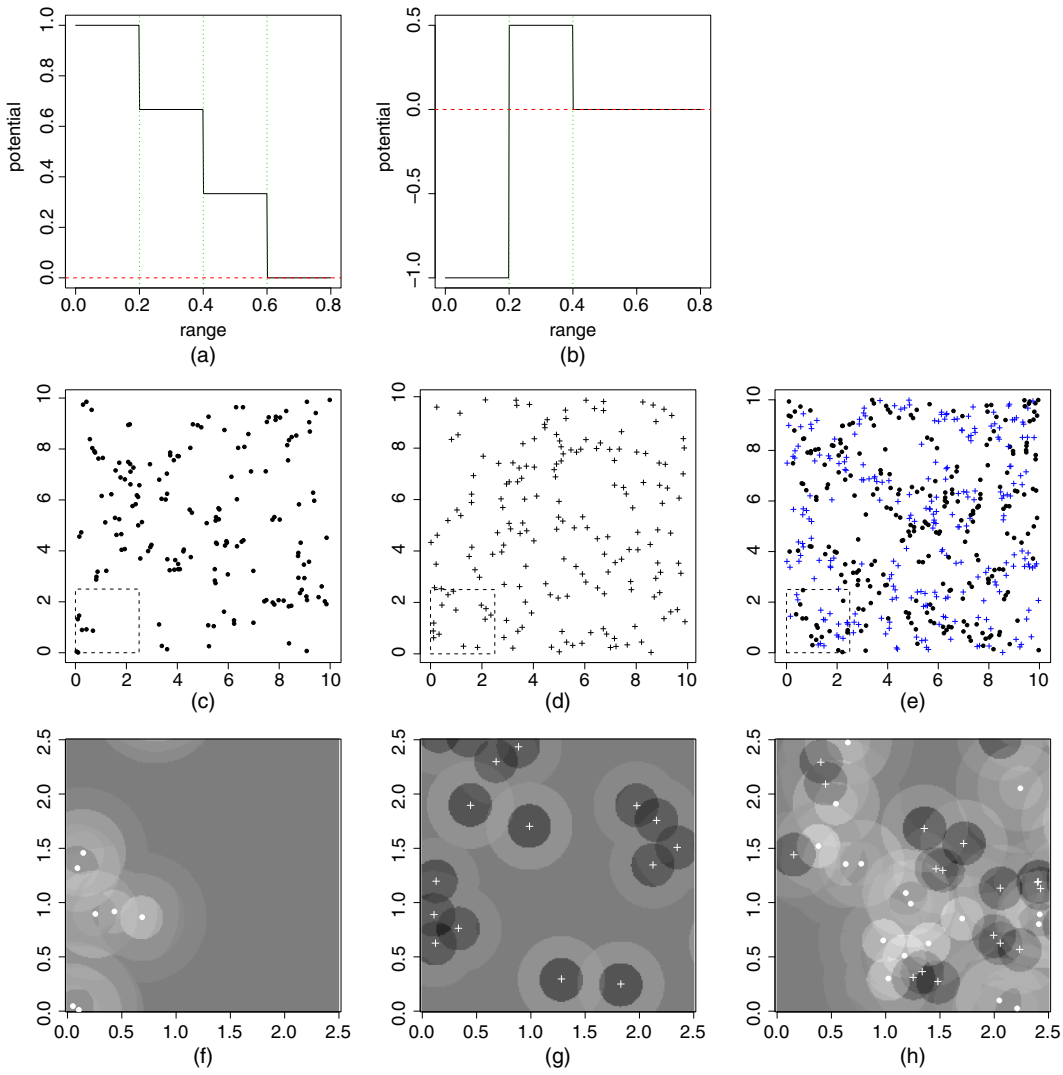

Fig. 12. Potentials (a) $\theta = (3, 2, 1)/3$ and (b) $\theta = (-1, 0.5)$ (:, range vector stops), simulations from the saturation model with $c = 1$ and potentials (c) $\theta = (3, 2, 1)/3$ and (d) $\theta = (-1, 0.5)$ and (e) a joint bivariate simulation with intrapotentials given by $\theta = (3, 2, 1)/3$ and interpotential given by $\theta = (-1, 0.5)$ and (f), (g), (h) conditional intensities of the saturation model given the simulations in (c), (d) and (e) respectively (in (h) just for type $i = 1$) (conditional intensities are on a log-scale; background grey colour means 0 or no potential; darker colour means negative and lighter colour means positive potential; in the corner $[0, 2.5]^2$)

Appendix B: Derivation of saturation approximation of the Papangelou terms under independence

We show how the probability of saturation of the components in the log-Papangelou conditional intensity can be derived by using the independence assumption. First we note that for a homogeneous Poisson process neighbour counts in any set depend on only the volume of the set. Thus to simplify from the annuli notation it suffices to consider only $r = r_k > r_{k-1} = 0$. Then

$$\omega(u, \mathbf{x}) = \min\{c, \mathrm{ne}(u, \mathbf{x})\} + \sum_{x \in \mathbf{x}} [\min\{c, \mathrm{ne}(x, \mathbf{x} \cup u)\} - \min\{c, \mathrm{ne}(x, \mathbf{x} \setminus u)\}] \tag{16}$$

$$= \min(c, \#[b(u, r) \cap \mathbf{x}]) + \sum_{x \in \mathbf{x}} \mathbf{1}\{x \in b(u, r)\} \mathbf{1}(\#[b(x, r) \cap \mathbf{x}] \in [0, c)) \tag{17}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function. Assume now that $\mathbf{x}$ comes from a Poisson point process with intensity λ. Write $a = \lambda |b(o, r)|$. The Georgii–Nguyen–Zessin formula (Illian *et al*. (2008), page 399) gives

$$\begin{aligned}
\mathbb{E}\{\omega(u, X)\} &= \mathbb{E}\{\min(c, \#[b(u, r) \cap X])\} + \mathbb{E}\Big[\sum_{x \in X} \mathbf{1}\{x \in b(u, r)\} \mathbf{1}\{\#[b(x, r) \cap \mathbf{x}] \in [0, c)\}\Big] \\
&= \mathbb{E}\{\min(c, \#[b(o, r) \cap X])\} + a\, \mathbb{E}[\mathbf{1}\{\#[b(o, r) \cap X] \in [0, c)\}].
\end{aligned}$$

Now, the random variable $y := \#[b(o, r) \cap X]$ is Poisson(a) distributed. Denote its cumulative distribution function with F_a. Then

$$\begin{aligned}
\mathbb{E}\{\omega(u, X)\} &= \sum_{k=0}^{c-1} P(y = k) k + c \sum_{k=c}^{\infty} P(y = k) + a \sum_{k=0}^{c-1} P(y = k) \\
&= a \sum_{l=0}^{c-2} P(y = l) + c\{1 - F_a(c-1)\} + a F_a(c-1) \\
&= c\{1 - F_a(c-1)\} + a\{F_a(c-1) + F_a(c-2)\},
\end{aligned}$$

giving the function $t(c)$ in the text. Note that we also use the fact that $\lambda \approx n/|W|$.

Appendix C: Illustration of the interaction and potential functions

Figs 12(a) and 12(b) depict two particular potential shapes with different $\mathbf{r}$s and βs, to illustrate their role. The first function corresponds to a decreasing attraction in range, and the second function has both a repulsion and an attraction component. How the attraction and repulsion impact the point pattern depends on the choice of the individual g_{ijk}s: in this example we use the saturation model $\mathbf{g}_{ij}$s with $c_{ijk} \equiv 1$. Figs 12(c) and 12(d) show univariate simulations from each of the two interaction functions. Fig. 12(e) shows a bivariate simulation with intratype interaction given by the first potential and intertype interaction given by the second potential. Figs 12(f)–12(h) show the conditional intensity (10) at the window locations $u \in W$ given the simulated patterns.

In the case of the bivariate pattern the conditional intensity is for points of the first type. The conditional intensities exhibit various features: for the first potential function, high potential locations are near data points, but only if the data points do not already have neighbours (those data points' neighbourhoods are already saturated). With the second potential any location too near the data points has a low potential (repulsion), but being too far from the data points is not encouraged by the potential either (attraction). In Fig. 12(h) we see the complex mixture of the first type's internal potential and the intratype potential: a point of the first type would be welcome near a point of its own kind (attraction) but unwelcome near a point of the second kind (repulsion).

References

Amburgey, T. L. (1986) Multivariate point process models in social research. *Socl Sci. Res.*, **15**, 190–207.

Andersen, I. and Hahn, U. (2016) Matérn thinned Cox processes. *Spatl Statist.*, **15**, 1–21.

Baddeley, A., Coeurjolly, J., Rubak, E. and Waagepetersen, R. (2014) Logistic regression for spatial Gibbs point processes. *Biometrika*, **101**, 377–392.

Baddeley, A. and van Lieshout, M. (1995) Area-interaction point processes. *Ann. Inst. Statist. Math.*, **47**, 601–619.

Baddeley, A. and Turner, R. (2000) Practical maximum pseudolikelihood for spatial point patterns. *Aust. New Zeal. J. Statist.*, **42**, 283–322.
Baddeley, A., Turner, R., Mateu, J. and Bevan, A. (2013) Hybrids of Gibbs point process models and their implementation. *J. Statist. Softwr.*, **55**, 1–43.
Baddeley, A., Turner, R., Møller, J. and Hazelton, M. (2004) Residual analysis for spatial point processes (with discussion). *J. R. Statist. Soc.* B, **67**, 617–666.
Baldeck, C. A., Harms, K. E., Yavitt, J. B., John, R., Turner, B. L., Valencia, R., Navarrete, H., Bunyavejchewin, S., Kiratiprayoon, S., Yaacob, A., Supardi, M. N. N., Davies, S. J., Hubbell, S. P., Chuyong, G. B., Kenfack, D., Thomas, D. W. and Dalling, J. W. (2013a) Habitat filtering across tree life stages in tropical forest communities. *Proc. R. Soc. Lond.* B, **280**, article 20130548.
Baldeck, C. A., Harms, K. E., Yavitt, J. B., John, R., Turner, B. L., Valencia, R., Navarrete, H., Davies, S. J., Chuyong, G. B., Kenfack, D., Thomas, D. W., Madawala, S., Gunatilleke, N., Gunatilleke, S., Bunyavejchewin, S., Kiratiprayoon, S., Yaacob, A., Supardi, M. N. N. and Dalling, J. W. (2013b) Soil resources and topography shape local tree community structure in tropical forests. *Proc. R. Soc. Lond.* B, **280**, article 20122532.
Breheny, P. and Huang, J. (2009) Penalized methods for bi-level variable selection. *Statist. Interfc.*, **2**, 369–380.
Brix, A. and Møller, J. (2001) Space-time multi type log Gaussian Cox processes with a view to modelling weeds. *Scand. J. Statist.*, **28**, 471–488.
Brown, C., Illian, J. B. and Burslem, D. F. R. P. (2016) Success of spatial statistics in determining underlying process in simulated plant communities. *J. Ecol.*, **104**, 160–172.
Chiu, S. N., Stoyan, D., Kendall, W. S. and Mecke, J. (2013) *Stochastic Geometry and Its Applications*. Chichester: Wiley.
Coeurjolly, J.-F. and Lavancier, F. (2013) Residuals and goodness-of-fit tests for stationary marked Gibbs point processes. *J. R. Statist. Soc.* B, **75**, 247–276.
Coeurjolly, J. F. and Rubak, E. (2013) Fast covariance estimation for innovations computed from a spatial Gibbs point process. *Scand. J. Statist.*, **40**, 669–684.
Condit, R. (1998) *Tropical Forest Census Plots*. Berlin: Springer.
Condit, R., Ashton, P. S., Manokaran, N., LaFrankie, J. V., Hubbell, S. P. and Foster, R. B. (1999) Dynamics of the forest communities at Pasoh and Barro Colorado: comparing two 50-ha plots. *Philos. Trans. R. Soc. Lond.* B, **354**, 1739–1748.
Diggle, P. J. and Milne, R. K. (1983) Bivariate Cox processes: some models for bivariate spatial point patterns. *J. R. Statist. Soc.* B, **45**, 11–21.
Diggle, P., Zheng, P. and Durr, P. (2005) Nonparametric estimation of spatial segregation in a multivariate point process: bovine tuberculosis in Cornwall, UK. *Appl. Statist.*, **54**, 645–658.
Flügge, A. J., Olhede, S. C. and Murrell, D. J. (2014) A method to detect subcommunities from multivariate spatial associations. *Meth. Ecol. Evoln*, **5**, 1214–1224.
Funwi-Gabga, N. and Mateu, J. (2012) Understanding the nesting spatial behaviour of gorillas in the Kagwene Sanctuary, Cameroon. *Stoch. Environ. Res. Risk Assessmnt*, **26**, 793–811.
Gates, D. and Westcott, M. (1986) Clustering estimates for spatial point distributions with unstable potentials. *Ann. Inst. Statist. Math.*, **38**, 123–135.
van der Geer, S. and Bühlmann, P. (2011) *Statistics for High-dimensional Data: Methods, Theory and Applications*. Berlin: Springer.
Gelfand, A. E., Schmidt, A. M., Banerjee, S. and Sirmans, C. (2004) Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, **13**, 263–312.
Gelfand, A., Schmidt, A. and Sirmans, C. (2002) Multivariate spatial process models: conditional and unconditional Bayesian approaches using coregionalization. *Technical Report*. Institute of Statistics and Decision Sciences, Duke University, Durham.
Geyer, C. (1999) Likelihood inference for spatial point processes. In *Stochastic Geometry: Likelihood and Computation* (eds O. E. Barndorff-Nielsen, W. S. Kendall and M. N. M. van Lieshout), pp. 79–140. London: Chapman and Hall.
Giraud, C. (2014) *Introduction to High-dimensional Statistics*. Boca Raton: Chapman and Hall–CRC.
Grabarnik, P. and Särkkä, A. (2001) Interacting neighbour point processes: some models for clustering. *J. Statist. Computn Simuln*, **68**, 103–125.
Grabarnik, P. and Särkkä, A. (2009) Modelling the spatial structure of forest stands by multivariate point processes with hierarchical interactions. *Ecol. Modllng*, **220**, 1232–1240.
Hastie, T., Tibshirani, R. and Friedman, J. (2001) *The Elements of Statistical Learning*. New York: Springer.
Högmander, H. and Särkkä, A. (1999) Multitype spatial point patterns with hierarchical interactions. *Biometrics*, **55**, 1051–1058.
Hubbell, S., Condit, R. and Foster, R. (2005) Barro Colorado forest census plot data. Smithsonian Tropical Research Institute, Panama City. (Available from `http://ctfs.si.edu/webatlas/datasets/bci`.)
Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008) *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester: Wiley.
Jaakkola, T. and Jordan, M. (2000) Bayesian parameter estimation via variational methods. *Statist. Comput.*, **10**, 25–37.

Kanagaraj, R., Wiegand, T., Comita, L. S. and Huth, A. (2011) Tropical tree species assemblages in topographical habitats change in time and with life stage. *J. Ecol.*, **99**, 1441–1452.
Lan, G., Getzin, S., Wiegand, T., Hu, Y., Xie, G., Zhu, H. and Cao, M. (2012) Spatial distribution and interspecific associations of tree species in a tropical seasonal rain forest of China. *PLOS One*, **7**, article e46074.
Lavancier, F. and Møller, J. (2016) Modelling aggregation on the large scale and regularity on the small scale in spatial point pattern datasets. *Scand. J. Statist.*, **43**, 587–609.
Law, R. and Dieckmann, U. (2000) Moment approximations of individual-based models. In *The Geometry of Ecological Interactions: Simplifying Spatial Complexity* (eds U. Dieckmann, R. Law and J. A. J. Metz), pp. 252–270. Cambridge: Cambridge University Press.
Law, R., Murrell, D. J. and Dieckmann, U. (2003) Population growth in space and time: spatial logistic equations. *Ecology*, **84**, 252–262.
van Lieshout, M. N. M. (2000) *Markov Point Processes and Their Applications*. London: Imperial College Press.
Meier, L., van de Geer, S. and Bühlmann, P. (2008) The group lasso for logistic regression. *J. R. Statist. Soc.* B, **70**, 53–71.
Mitchell, T. J. and Beauchamp, J. J. (1988) Bayesian variable selection in linear regression. *J. Am. Statist. Ass.*, **83**, 1023–1032.
Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P. and Tita, G. E. (2011) Self-exciting point process modeling of crime. *J. Am. Statist. Ass.*, **106**, 100–108.
Morueta-Holme, N., Blonder, B., Sandel, B., Mcgill, B. J., Peet, R. K., Ott, J. E., Violle, C., Enquist, B. J., Jørgensen, P. M. and Svenning, J. C. (2016) A network approach for inferring species associations from co-occurrence data. *Ecography*, **39**, 1139–1150.
Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H. and Hahn, U. (2017) Global envelope tests for spatial processes. *J. R. Statist. Soc.* B, **79**, 381–404.
Olsbo, V., Myllymäki, M., Waller, L. A. and Särkkä, A. (2013) Development and evaluation of spatial point process models for epidermal nerve fibers. *Math. Biosci.*, **243**, 178–189.
Ormerod, J. and Wand, M. (2010) Explaining variational approximations. *Am. Statistn*, **64**, 140–153.
Punchi-Manage, R., Getzin, S., Wiegand, T., Kanagaraj, R., Gunatilleke, S. C., Gunatilleke, I. N., Wiegand, K. and Huth, A. (2013) Effects of topography on structuring local species assemblages in a Sri Lankan mixed dipterocarp forest. *J. Ecol.*, **101**, 149–160.
R Core Team (2017) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Scheipl, F. (2011) spikeSlabGAM: Bayesian variable selection, model choice and regularization for generalized additive mixed models in R. *J. Statist. Softwr.*, **43**, 1–24.
Schoenberg, F. P. (2003) Multidimensional residual analysis of point process models for earthquake occurrences. *J. Am. Statist. Ass.*, **98**, 789–795.
Schreeg, L. A., Kress, W. J., Erickson, D. L. and Swenson, N. G. (2010) Phylogenetic analysis of local-scale tree soil associations in a lowland moist tropical forest. *PLOS One*, **5**, article e13685.
Shimatani, K. (2001) Multivariate point processes and spatial variation of species diversity. *Forst Ecol. Mangmnt*, **142**, 215–229.
Simon, N., Friedman, J., Hastie, T. and Tibshirani, R. (2013) A sparse-group lasso. *J. Computnl Graph. Statist.*, **22**, 231–245.
Stoica, R. S., Martínez, V. J. and Saar, E. (2007) A three-dimensional object point process for detection of cosmic filaments. *Appl. Statist.*, **56**, 459–477.
Stoyan, D. (1979) Interrupted point processes. *Biometr. J.*, **21**, 607–610.
Tibshirani, R. (1996) Regression selection and shrinkage via the lasso. *J. R. Statist. Soc.* B, **58**, 267–288.
Uriarte, M., Condit, R., Canham, C. D. and Hubbell, S. P. (2004) A spatially explicit model of sapling growth in a tropical forest: does the identity of neighbors matter. *J. Ecol.*, **92**, 348–360.
Velázquez, E., Martínez, I., Getzin, S., Moloney, K. A. and Wiegand, T. (2016) An evaluation of the state of spatial point pattern analysis in ecology. *Ecography*, **39**, 1042–1055.
Waagepetersen, R., Guan, Y., Jalilian, A. and Mateu, J. (2016) Analysis of multispecies point patterns by using multivariate log-Gaussian Cox processes. *Appl. Statist.*, **65**, 77–96.
Wiegand, T., Huth, A., Getzin, S., Wang, X., Hao, Z., Gunatilleke, C. V. S. and Gunatilleke, I. A. U. N. (2012) Testing the independent species' arrangement assertion made by theories of stochastic geometry of biodiversity. *Proc. R. Soc. Lond.* B, **279**, 3312–3320.
Yang, Q.-S., Shen, G.-C., Liu, H.-M., Wang, Z.-H., Ma, Z.-P., Fang, X.-F., Zhang, J. and Wang, X.-H. (2016) Detangling the effects of environmental filtering and dispersal limitation on aggregated distributions of tree and shrub species: life stage matters. *PLOS One*, **11**, article e0156326.
Yuan, M. and Lin, Y. (2006) Model selection and estimation in regression with grouped variables. *J. R. Statist. Soc.* B, **68**, 49–67.
Yue, Y. and Loh, J. M. (2014) Variable selection for inhomogeneous spatial point process models. *Can. J. Statist.*, **42**, 579–596.