Biometrika (1998), **85**, 2, *pp.* 251–267
Printed in Great Britain

Poisson/gamma random field models for spatial statistics

BY ROBERT L. WOLPERT
Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina 27708-0251, U.S.A.
wolpert@stat.duke.edu

AND KATJA ICKSTADT
Department of Statistics, University of North Carolina, Chapel Hill, North Carolina 27599-3260, U.S.A.
katja@stat.unc.edu

SUMMARY

Doubly stochastic Bayesian hierarchical models are introduced to account for uncertainty and spatial variation in the underlying intensity measure for point process models. Inhomogeneous gamma process random fields and, more generally, Markov random fields with infinitely divisible distributions are used to construct positively autocorrelated intensity measures for spatial Poisson point processes; these in turn are used to model the number and location of individual events. A data augmentation scheme and Markov chain Monte Carlo numerical methods are employed to generate samples from Bayesian posterior and predictive distributions. The methods are developed in both continuous and discrete settings, and are applied to a problem in forest ecology.

Some key words: Bayesian mixture model; Bioabundance; Cox process; Data augmentation; Lévy process; Markov chain Monte Carlo; Simulation.

1. INTRODUCTION

Having observed counts and some collateral information in some region we would like to make inferences about the probability mechanism underlying these counts, exploring their dependence and relation to covariates. In many applications the data will include spatial information, such as the exact locations of the observations. Statistical methods and models that reflect spatial variation explicitly are used commonly now in many areas of applied statistics, including geostatistics, disease mapping and image restoration; see Cressie (1993) or Diggle (1983) for excellent overviews. In this paper we present a new class of models for exploring the possibility that counts from nearby locations are similar, in being positively correlated or exhibiting spatial dependence in some other way.

A number of modelling approaches have been proposed for studying this question. Positive association is exhibited by inhomogeneous Poisson processes with continuously varying intensity, including doubly-stochastic or Cox processes and Poisson cluster models (Cox, 1955; Cox & Isham, 1980, Ch. 6; Neyman & Scott, 1958). Although some common Markov point processes (Strauss, 1975) are, like the auto-Poisson processes of Besag (1974), purely inhibitory (Kelly & Ripley, 1976), others do permit positive association (Baddeley & van Lieshout, 1995; Ripley, 1977; Ripley & Kelly, 1977).

Perhaps the most common approach is to model either the counts themselves, after a variance-stabilising transformation, or the intensity for an underlying Poisson model, after a logarithmic transformation, as a Gaussian Markov random field with a partially uncertain covariance structure (Besag, 1974; Cressie, 1993, p. 402; Cressie & Chan, 1989; Møller, Syversveen & Waagepetersen, 1998). Unfortunately, Gaussian models fail to reflect the counts' discrete nature, while a log-normal approach (Bernardinelli et al., 1997; Besag, York & Mollié, 1991; Clayton & Kaldor, 1987) does not scale properly under aggregation and refinement of partitions: the logarithmic structure leads to products rather than sums for the Poisson means for unions of neighbouring sets.

Traditionally, inference in spatial statistics has been based on a combination of ad hoc nonparametric techniques, such as distance-based methods and second-order methods, kernel smoothing (Silverman, 1986, Ch. 4), maximum-likelihood parameter estimates and associated clever approximations and simulation-based estimation and testing. The emergence of Markov chain Monte Carlo algorithms (Tierney, 1994; Gilks, Richardson & Spiegelhalter, 1996) has made it practicable, and increasingly common, to apply Bayesian statistical methods to problems in spatial statistics, particularly for Markov random field models where the complete conditional distributions required for the Gibbs sampling approach are usually available (Gelfand & Smith, 1990).

In this paper we introduce a natural and flexible class of Bayesian hierarchical models for analysing spatially dependent count data without any arbitrary discretisation of the data. The models are doubly stochastic Poisson processes whose intensities are convolutions or mixtures of inhomogeneous, infinitely divisible random fields; as such they generalise the conjugate Poisson/gamma hierarchical models of Clayton & Kaldor (1987), and may be viewed as limits in probability of Poisson cluster processes. The present methods extend, and in special cases reduce to, those based on mixtures of Dirichlet processes (Antoniak, 1974; Ferguson, 1973). Inference is based on simulations from the joint posterior distribution of all quantities of interest.

At the first, that is the lowest, stage of the hierarchy the number of points in each set A in some region $\mathcal{X}$ of Euclidean space $\mathbb{R}^d$ is modelled as a Poisson-distributed random variable $N(A)$, whose mean $\lambda(A)$ is the unobserved value of a random measure $\Lambda(A)$.

Doubly stochastic models in which the 'increments' $\Lambda(A_i)$ assigned to disjoint sets A_i are independent, such as the gamma process of Ferguson (1974), would necessarily feature independent counts $N(A_i)$ as well, and so could not be used to study possible spatial dependence. Instead at the second stage of the hierarchy we construct the intensity measure as a kernel mixture $\Lambda(A) \equiv \int_{\mathcal{S}} k(A, s)\Gamma(ds)$ for an independent-increment infinitely divisible random measure $\Gamma(ds)$ on an auxiliary space $\mathcal{S}$. In our examples in § 5, $\mathcal{S}$ is a subset of Euclidean space or a discrete set. For exposition we take $\Gamma(ds)$ to have the gamma process distribution $\Gamma(ds) \sim \mathrm{Ga}\{\alpha(ds), \beta(s)^{-1}\}$, with shape measure $\alpha(ds)$ and inverse scale function $\beta(s) > 0$.

To express uncertainty about the kernel and some aspects of the distribution of $\Gamma(ds)$ within the Bayesian paradigm we introduce at the third stage of the hierarchy a prior probability distribution $\pi(d\theta)$ on a parameter space Θ, and indicate by superscripts the dependence on $\theta \in \Theta$ of $k(x, s) = k^\theta(x, s)$, $\alpha(ds) = \alpha^\theta(ds)$ and $\beta(s) = \beta^\theta(s)$. Interest centres on the uncertain kernel $k^\theta(x, s)$, on the unobserved Poisson mean $\Lambda(x)$, on features of the distributions or even the specific values of the underlying process $\Gamma(ds)$, or on derived quantities such as linear trends and other measures of systematic spatial variation.

Section 2 introduces a new and very efficient scheme, the Inverse Lévy Measure algorithm, for sampling from inhomogeneous gamma random fields and other random measures

assigning independent infinitely divisible random variables to disjoint sets. We believe that this algorithm, based on ideas of Lévy (1937) and generalising more recent work of several authors (Bondesson, 1982; Ferguson & Klass, 1972; Laud, Smith & Damien, 1996) to our inhomogeneous spatial setting, will be useful in a variety of other problems. In § 3 the Poisson/gamma random field models are introduced, and in § 4 the Inverse Lévy Measure algorithm becomes the key to implementing a hybrid Metropolis/Gibbs approach with data augmentation (Tanner & Wong, 1987) to evaluate posterior distributions.

In § 5 the methods are illustrated with an example from forest ecology, studying the spatial distribution of hickory trees in a portion of Duke Forest in Durham, North Carolina. We analyse the data twice, first in a continuous setting and then in a discrete one. The discrete form is appropriate for applications in which census counts for subregions are given instead of exact locations; this is typical in disease mapping, for example, where data are customarily aggregated and reported by hospital, county or state to preserve confidentiality. The methods are then discussed in § 6. All but one-line proofs are deferred to an Appendix.

2. Poisson and gamma random fields

The simplest way to specify or identify the probability distribution of nonnegative random measures such as $N(A)$ is to give the moment generating function for the associated random field $N[\phi] \equiv \int_{\mathscr{X}} \phi(x) N(dx)$ defined in the obvious way by $N[\phi] \equiv \sum \phi_i N(A_i)$ for nonnegative simple functions $\phi(x) = \sum \phi_i 1_{A_i}(x)$, where $1_A(x)$ denotes the indicator function equal to one if $x \in A$ and zero otherwise, and extended by continuity to a wider class. This is usually done by giving the Laplace exponent, that is the negative logarithm $\mathscr{L}_N(\phi) \equiv -\log[E\{\exp(-N[\phi])\}]$ (Karr, 1991, p. 9).

For example, the distributions of random measures assigning independent Poisson random variables $N(A) \sim \mathrm{Po}\{\lambda(A)\}$, with mean $\lambda(A)$ for some measure $\lambda(dx)$ on $\mathscr{X}$, and gamma random variables $\Gamma(B) \sim \mathrm{Ga}\{\alpha(B), \beta^{-1}\}$, with inverse scale parameter $\beta > 0$ and shape measure $\alpha(ds)$ on some set $\mathscr{S}$, are determined by the Laplace exponents

$$\mathscr{L}_N(\phi) = \int_{\mathscr{X}} (1 - e^{-\phi(x)}) \lambda(dx), \quad \mathscr{L}_\Gamma(\phi) = \int_{\mathscr{S}} \log\{1 + \phi(s)/\beta\} \alpha(ds). \tag{2·1}$$

From Lévy (1937, Ch. VII, § 53–55) and Jacod & Shiryaev (1987, Ch. II, § 4c) it follows that any random field, including the Poisson and gamma fields, that assigns independent infinitely divisible random variables to disjoint sets and satisfies appropriate regularity conditions must have a Laplace exponent of the form $\mathscr{L}(\phi) = \int_{\mathscr{S}} (1 - e^{-u\phi(s)}) \nu(du, ds)$, generalising that of the Poisson. The gamma random field, for example, has 'Lévy measure' $\nu_\Gamma(du, ds) = e^{-u\beta} u^{-1}\, du\alpha(ds)$. This has the following two important consequences for us.

(i) It suggests that the gamma process can be extended to nonconstant $\beta(s)$, with Laplace exponent $\mathscr{L}_\Gamma(\phi) = \int_{\mathscr{S}} \log\{1 + \phi(s)/\beta(s)\} \alpha(ds)$ and Lévy measure $\nu_\Gamma(du, ds) = e^{-u\beta(s)} u^{-1}\, du\alpha(ds)$. This process was introduced by Dykstra & Laud (1981), who called it the Extended Gamma process, for the special case of $\mathscr{S} = \mathbb{R}_+$; an explicit but approximate construction appears in Laud et al. (1996) under the restriction of right-continuity for $\beta(s)$. We will see below that the restrictions of $\mathscr{S}$ and $\beta(s)$ are unnecessary.

(ii) It suggests that the gamma process can be constructed from a Poisson process

$H(du, ds)$ on $\mathbb{R}_+ \times \mathcal{S}$ with mean $E\{H(du, ds)\} = \nu(du, ds)$ by setting

$$\Gamma[\phi] = \int\int_{\mathbb{R}_+ \times \mathcal{S}} u\phi(s) H(du, ds)$$

or, equivalently,

$$\Gamma[\phi] = \sum_{m<\infty} v_m \phi(\sigma_m) \quad \text{or} \quad \Gamma(A) = \sum_{\sigma_m \in A} v_m, \tag{2·2}$$

where the sums extend over the random countable support $\{(v_m, \sigma_m)\}$ of $H(du, ds)$.

Both suggestions (i) and (ii) above are borne out by the following result, in which $E_1(t) \equiv \int_t^\infty e^{-u} u^{-1}\, du$ denotes the exponential integral function (Abramowitz & Stegun, 1964, p. 228).

THEOREM 1. *Let $\alpha(s) \geqslant 0$ and $\beta(s) > 0$ be measurable functions on a space $\mathcal{S}$. Let $\{\sigma_m\}$ be independent identically distributed draws from any probability distribution $\Pi(ds)$ on $\mathcal{S}$, and let $\tau_m \geqslant 0$ be the successive jump times of a standard Poisson process. Set $\tau(u, s) \equiv E_1\{u\beta(s)\}\alpha(s)$ and $v_m \equiv \sup\{u \geqslant 0 : \tau(u, \sigma_m) \geqslant \tau_m\}$, that is,*

$$v_m \equiv E_1^{-1}\{\tau_m / \alpha(\sigma_m)\}\beta(\sigma_m)^{-1}, \tag{2·3}$$

or $v_m \equiv 0$ if $\alpha(\sigma_m) = 0$. Then the random field defined by $\Gamma[\phi] \equiv \sum_{m<\infty} v_m \phi(\sigma_m)$ for bounded measurable $\phi(s)$ has the gamma process distribution $\Gamma(ds) \sim \mathrm{Ga}\{\alpha(ds), \beta(s)^{-1}\}$ for the measure $\alpha(ds) \equiv \alpha(s)\Pi(ds)$.

COROLLARY 1. *The gamma random field $\Gamma(ds) \sim \mathrm{Ga}\{\alpha(ds), \beta(s)^{-1}\}$ can be approximated in distribution to arbitrarily high accuracy by the following.*

INVERSE LÉVY MEASURE ALGORITHM

Step 1. Fix a large integer M and choose any convenient distribution $\Pi(ds)$ on $\mathcal{S}$ from which it is easy to draw samples and such that the shape measure $\alpha(ds)$ has a density $\alpha(s) \equiv \alpha(ds)/\Pi(ds)$; see guidance on choices below.

Step 2. Generate M independent identically distributed draws $\{\sigma_m\}_{m \leqslant M}$ from $\Pi(ds)$.

Step 3. Generate the first M jumps $\{\tau_m\}_{m \leqslant M}$ of a standard Poisson process, for example by adding successive independent exponential random variables.

Step 4. Set $v_m \equiv E_1^{-1}\{\tau_m / \alpha(\sigma_m)\}\beta(\sigma_m)^{-1}$.

Step 5. Set $\Gamma_M(ds) \equiv \sum_{m \leqslant M} v_m \delta_{\sigma_m}(ds) \doteq \Gamma(ds)$, $\Gamma_M[\phi] \equiv \sum_{m \leqslant M} v_m \phi(\sigma_m) \doteq \Gamma[\phi]$.

COROLLARY 2. *The joint distribution of the jump sizes and locations*

$$\{(v_m, \sigma_m)\}_{m \leqslant M} \subset \mathbb{R}_+ \times \mathcal{S}$$

in the Inverse Lévy Measure algorithm has a density with respect to the product measure $\prod_{m \leqslant M} dv_m \Pi(d\sigma_m)$ proportional to

$$e^{-E_1\{v_M \beta(\sigma_M)\}\alpha(\sigma_M)} \prod_{m \leqslant M} \{v_m^{-1} e^{-v_m \beta(\sigma_m)} \alpha(\sigma_m)\}$$

on the set of $\{(v_m, \sigma_m)\}_{m \leqslant M}$ for which the function

$$\tau(u, s) \equiv E_1\{u\beta(s)\}\alpha(s)$$

satisfies $\tau(v_m, \sigma_m) < \tau(u_{m+1}, \sigma_{m+1})$ for every $m < M$.

Proof. This requires simply a change of variables (2·3) from $\tau_m = \tau(v_m, \sigma_m)$ to v_m. □

Remark 1. The probability distribution Π can be written uniquely as the sum of continuous and discrete parts $\Pi(ds) = \Pi_c(ds) + \sum \pi_n \delta_{s_n}(ds)$, leading to similar representations for the shape measure

$$\alpha(ds) = \alpha(s)\Pi_c(ds) + \sum \pi_n \alpha(s_n)\delta_{s_n}(ds)$$

and random field $\Gamma(ds) = \Gamma_c(ds) + \sum v_n \delta_{s_n}(ds)$, with $v_n \sim \mathrm{Ga}\{\pi_n \alpha(s_n), \beta(s_n)^{-1}\}$ independent of the gamma random field $\Gamma_c \sim \mathrm{Ga}\{\alpha(s)\Pi_c(ds), \beta(s)^{-1}\}$; often it is more efficient to use the Inverse Lévy Measure algorithm for Γ_c while sampling directly the independent random variables v_n from gamma distributions. Upon relabelling we recover representation (2·2).

The remarkable feature of Theorem 1 is the freedom to choose any convenient sampling distribution $\Pi(ds)$ that dominates $\alpha(ds)$, and any measurable function $\beta(s) > 0$. The Inverse Lévy Measure algorithm compensates for oversampling, where $\alpha(s) = \alpha(ds)/\Pi(ds)$ is small, or undersampling, where $\alpha(s)$ is large, by generating small or large jumps v_m, respectively. Naturally some choices of $\Pi(ds)$ will sample more efficiently than others, so that $\Gamma(ds)$ will be better approximated by the partial sum $\Gamma_M(ds)$ for modest M.

Appropriate choices for M and $\Pi(ds)$ can be guided by considering the expected truncation error. From the estimates $E_1(z) = -\gamma - \log(z) + O(z)$ (Abramowitz & Stegun, 1964, § 5.1.11) and $\tau_M \doteq M$ we find that $v_m \doteq \exp\{-\gamma - m/\alpha(\sigma_m)\}/\beta(\sigma_m)$; for nearly constant $\alpha(s) \doteq \alpha$, that is, for $\Pi(ds)$ nearly proportional to $\alpha(ds)$, and $\beta(s) \doteq \beta$, this is approximately a geometric series, with relative truncation error about $\exp(-M/\alpha)$. For $\alpha(\mathscr{S}) < \infty$ and constant $\beta(s) \equiv \beta$ the choice $\Pi(ds) = \alpha(ds)/\alpha(\mathscr{S})$ makes $\alpha(s) = \alpha(\mathscr{S})$ constant too, and the Inverse Lévy Measure algorithm will generate the v_m in strictly decreasing order starting at the largest mass point of $\Gamma(ds)$. In this case one might prefer to select a small $\varepsilon > 0$ and sample a random number M_ε of times until ε exceeds the smallest jump included and hence exceeds each of the jumps omitted, to get an expected truncation error that obeys the simple bound $E\{\Gamma(A) - \Gamma_{M_\varepsilon}(A)\} \leqslant \varepsilon\alpha(A)$. In this special case, in one dimension, $\mathscr{S} = (0, 1) \subset \mathbb{R}^1$, with Lebesgue measure for $\alpha(ds)$ and $\Pi(ds)$, the present method reduces to that of Bondesson (1982) and generalises the method of Ferguson & Klass (1972).

Our simulation is approximate only in that we cannot draw infinitely many points (v_m, σ_m); the points we do draw are from exactly the right distribution, without approximation. Other published methods for simulating draws from some of these distributions (Damien, Laud & Smith, 1995; Laud et al., 1996) do entail approximations and moreover impose the restrictions that $\mathscr{S}$ be one-dimensional and that the Lévy measure $\nu(du, ds)$ have a product density with respect to Lebesgue measure $du\,ds$. Also, since they do not generate the largest jumps first, they require larger samples to ensure the same fraction of the total mass $\Gamma(\mathscr{S})$.

The Inverse Lévy Measure algorithm extends easily to any other infinitely divisible independent-increment process with Lévy measure $\nu(du, ds) = \nu(u, s)\,du\Pi(ds)$, simply by replacing $\tau(u, s) \equiv E_1\{u\beta(s)\}\alpha(s)$ with $\tau(u, s) \equiv \int_u^\infty \nu(v, s)\,dv$; hence the name 'Inverse Lévy Measure'. The one-sided stable process $S[\phi]$ of index $\xi(s) \in (0, 1)$ and intensity measure $\alpha(ds) = \alpha(s)\Pi(ds)$, for example, has Lévy measure $\nu(du, ds) = \xi(s)u^{-1-\xi(s)}\,du\alpha(ds)$ and so $\tau(u, s) = u^{-\xi(s)}\alpha(s)$ leads to a sampling scheme with $v_m = \{\alpha(\sigma_m)/\tau_m\}^{1/\xi(\sigma_m)}$; note that permitting the stable index $\xi(s)$ to be nonconstant entails no additional difficulty.

3. Models

3·1. *The Poisson/gamma model*

At the top, that is the third, level of the hierarchy let $\pi(d\theta)$ be a probability distribution on a Borel measurable space Θ.

At the middle, second stage of the hierarchy, given $\theta \in \Theta$, let $\Gamma(ds)$ be the gamma random field $\Gamma(ds) \sim \mathrm{Ga}\{\alpha^\theta(ds), \beta^\theta(s)^{-1}\}$ constructed in Theorem 1. Define an integrable function $\Lambda(x)$ on a measure space $\{\mathscr{X}, w(dx)\}$ by $\Lambda(x) \equiv \int_{\mathscr{S}} k^\theta(x, s)\Gamma(ds)$ for some nonnegative integral kernel $k^\theta(x, s)$ satisfying the integrability condition

$$\iiint_{\mathscr{X}\times\mathscr{S}\times\Theta} k^\theta(x, s)w(dx)\beta^\theta(s)^{-1}\alpha^\theta(ds)\pi(d\theta) < \infty;$$

this condition ensures that the random measure $\Lambda(dx) \equiv \Lambda(x)w(dx)$ and integral kernel $k^\theta(dx, s) \equiv k^\theta(x, s)w(dx)$ will be well defined.

At the bottom, first stage of the hierarchy, given θ and $\Gamma(ds)$, take $N(dx)$ to be a Poisson measure with conditional mean $\Lambda(dx)$, assigning conditionally independent $\mathrm{Po}\{\Lambda(A_i)\}$ random variables $N(A_i)$ to disjoint Borel sets $A_i \subset \mathscr{X}$; the unconditional dependence of the $\{N(A_i)\}$ arises from that of the $\{\Lambda(A_i)\}$. Multiple points are possible if $\Lambda(dx)$ has atoms; see the example in § 5·3. Thus $N(dx)$ is modelled as a doubly stochastic Poisson process with hierarchical representation:

$$\begin{aligned} \theta \sim \pi(d\theta), \quad \Gamma(ds)|_\theta \sim \mathrm{Ga}\{\alpha^\theta(ds), \beta^\theta(s)^{-1}\}, \\ \Lambda(dx) \equiv \int_{\mathscr{S}} k^\theta(dx, s)\Gamma(ds), \quad N(dx)|_{\theta,\Gamma} \sim \mathrm{Po}\{\Lambda(dx)\}. \end{aligned} \tag{3·1}$$

The representation, which depends upon $k^\theta(x, s)$ and $\beta^\theta(s)$ only through their ratio $k^\theta(x, s)/\beta^\theta(s)$, could be made unique by requiring $\beta^\theta(s) \equiv 1$ or $k^\theta(\mathscr{X}, s) \equiv 1$, but the present over-parameterisation facilitates Bayesian updating, which affects only $\beta^\theta(s)$, see § 4, and the inclusion of covariates, which affects only $k^\theta(x, s)$, see § 6.

The conditional mean and covariance of $N(dx)$, given θ, are easily computed:

$$\begin{aligned} E^\theta\{N(A)\} &= \int_{\mathscr{S}} k^\theta(A, s)\beta^\theta(s)^{-1}\alpha^\theta(ds), \\ \mathrm{cov}^\theta\{N(A), N(B)\} &= \int_{\mathscr{S}} k^\theta(A\cap B, s)\beta^\theta(s)^{-1}\alpha^\theta(ds) + \int_{\mathscr{S}} k^\theta(A, s)k^\theta(B, s)\beta^\theta(s)^{-2}\alpha^\theta(ds). \end{aligned}$$

We now seek $k^\theta(x, s)$, $\alpha^\theta(ds)$ and $\beta^\theta(s)$ flexible enough to approximate the anticipated point process means and covariances in an illustrative example.

3·2. *An illustration*

In § 5 we study an example from forest ecology in which $N(dx)$ counts the number and locations of hickory trees at least 10 m from the boundary of the 140 m × 140 m Bormann research plot in Duke Forest in Durham, North Carolina. We represent these regions by

$$\mathscr{X} \equiv [10, 130] \times [10, 130] \subset \mathscr{S} \equiv [0, 140] \times [0, 140] \subset \mathbb{R}^2.$$

For $\theta \in \mathbb{R}^2 \times \mathbb{R}_+$ let $\Gamma(ds)$ be a gamma random field with uniform shape measure $\alpha^\theta(ds) \equiv e^{\theta_1}\,ds$ and constant scale $\beta^\theta(s)^{-1} \equiv e^{\theta_2}$. Let $k^\theta(x, s)$ be a Gaussian kernel density $k^\theta(x, s) \equiv (\pi\theta_3^2)^{-1}\exp(-|x-s|^2/\theta_3^2)$ with respect to the reference measure $w(dx) \equiv 10^{-4}\,dx$ on $\mathscr{X}$, area measured in hectares, normalised by the condition $\int_{\mathbb{R}^2} k^\theta(x, s)\,ds \equiv 1$, supporting

a range of possible correlation distances. With these choices the approximate means and variances become

$$E^{\theta}\{N(A)\} = \int_{\mathscr{S}} k^{\theta}(A, s)\beta^{\theta}(s)^{-1}\alpha^{\theta}(ds) \simeq \int_{A} e^{\theta_1+\theta_2} w(dx) = e^{\theta_1+\theta_2} w(A),$$

$$\operatorname{var}^{\theta}\{N(A)\} \simeq E^{\theta}\{N(A)\} + \iint_{A\times A} (2\pi\theta_3^2)^{-1} e^{\theta_1+2\theta_2-|x-y|^2/2\theta_3^2} w(dx)w(dy)$$

$$\simeq \begin{cases} e^{\theta_1+\theta_2} w(A)[1 + \{e^{\theta_2}/(2\pi\theta_3^2)\}w(A)] & \text{if } \operatorname{diam}(A) \ll \theta_3, \\ e^{\theta_1+\theta_2} w(A)(1 + 10^{-4}e^{\theta_2}) & \text{if } \operatorname{diam}(A) \gg \theta_3. \end{cases}$$

The first two approximations are quite close for $A \subset \mathscr{X}$ whenever $\theta_3 \ll 10$ m and would be exact for $\mathscr{S} = \mathscr{X} = \mathbb{R}^2$, where the spatial point process $N(dx)$ would be stationary and isotropic with intensity $\lambda = e^{\theta_1+\theta_2}10^{-4}$and Ripley's function

$$K(t) = \pi t^2 + e^{-\theta_1}(1 - e^{-t^2/2\theta_3^2})$$

(Diggle, 1983, § 4.1). The expected number of trees within distance t of an arbitrary point and of an arbitrary tree would then be $\lambda\pi t^2$ and

$$\lambda K(t) = \lambda\pi t^2 + 10^{-4}e^{\theta_2}(1 - e^{-t^2/2\theta_3^2}),$$

respectively.

Evidently $e^{\theta_1+\theta_2}$ is the mean density, in trees per hectare; $10^{-4}e^{\theta_2}$ is the large-scale overdispersion or, alternatively, the excess number of trees close to a randomly selected tree; and θ_3 is a measure of the distance within $\mathscr{X}$ over which any spatial interaction extends. The overdispersion is smaller for small sets A, and is nearly proportional to $w(A)$. A prior distribution $\pi(d\theta)$ is selected in § 5 to express prior belief about these features. For future use note that $k^{\theta}(\mathscr{X}, s) \equiv \int_{\mathscr{X}} k^{\theta}(x, s)w(dx)$ in this example may be written in terms of the standard normal distribution function as

$$k^{\theta}(\mathscr{X}, s) = \left\{\Phi\left(\frac{130 - s_1}{\theta_3/2^{\frac{1}{2}}}\right) - \Phi\left(\frac{10 - s_1}{\theta_3/2^{\frac{1}{2}}}\right)\right\}\left\{\Phi\left(\frac{130 - s_2}{\theta_3/2^{\frac{1}{2}}}\right) - \Phi\left(\frac{10 - s_2}{\theta_3/2^{\frac{1}{2}}}\right)\right\} 10^{-4}, \tag{3·2}$$

which, if θ_3 is small, is approximately equal to 10^{-4} for s in the interior of $\mathscr{X}$ but near zero for s far from $\mathscr{X}$; in particular it is a nonconstant function of $s = (s_1, s_2) \in \mathscr{S}$.

4. The computational algorithm

4·1. *Introduction*

An exact likelihood function for θ may be evaluated as a convolution from the hierarchical mixture representation (3·1), but the computations are tedious and numerically unstable. Instead we follow a Markov chain Monte Carlo computational approach based on (3·1) to evaluate the joint posterior distribution of the uncertain quantities θ and $\Gamma = \Gamma(ds)$ and functions thereof by simulating steps (θ^t, Γ^t) from an ergodic Markov chain constructed to have as a stationary distribution the posterior distribution of θ and Γ given the observed count data $N(dx)$. The key computational requirements are the complete

conditional distributions of each uncertain quantity given all the others, evaluated as follows.

4·2. *Augmentation and the complete conditional distributions*

First consider the conditional distribution of Γ given θ and $N(dx)$. The point distribution $N(dx)$ is a finite integer-valued measure on $\mathscr{X}$, and so can be represented as the sum of a random number $N_{\mathscr{X}} \sim \mathrm{Po}\{\Lambda(\mathscr{X})\}$ of unit point masses at not-necessarily-distinct points $X_n = x_n$, drawn conditionally independently from the mixture distribution

$$X_n|_{\theta,\Gamma,N_{\mathscr{X}}} \sim \Lambda(dx)/\Lambda(\mathscr{X}) = \int_{\mathscr{S}} k^\theta(dx, s)\Gamma(ds)/\Lambda(\mathscr{X}).$$

For each $n \leqslant N_{\mathscr{X}}$ resolve the mixture by drawing an auxiliary random variable $S_n = s_n \in \mathscr{S}$ from the distribution $S_n|_{\theta,\Gamma,N} \sim k^\theta(x_n, s_n)\Gamma(ds_n)/\Lambda(x_n)$, and let

$$Z(dx\,ds) \equiv \textstyle\sum_{n \leqslant N_{\mathscr{X}}} \delta_{(x_n, s_n)}(dx\,ds)$$

be the random measure on $\mathscr{X} \times \mathscr{S}$ assigning unit mass to each pair (x_n, s_n). Of course the unaugmented data may be recovered as the first marginal measure $N(dx) = Z_1(dx) \equiv Z(dx \times \mathscr{S})$, while the augmentation points $\{S_n = s_n\}$ can be recovered as the second marginal, $Z_2(ds) \equiv Z(\mathscr{X} \times ds)$. With this data augmentation we have the following conjugacy property.

THEOREM 2. *The conditional distribution of* $\Gamma(ds)$, *given* θ *and* $Z(dx\,ds)$, *is given by*

$$\Gamma(ds)|_{\theta,Z} \sim \mathrm{Ga}[(\alpha^\theta + Z_2)(ds),\ \{\beta^\theta(s) + k^\theta(\mathscr{X}, s)\}^{-1}]. \tag{4·1}$$

Note that, even if we begin with a constant $\beta^\theta(s) \equiv \beta^\theta$, we are forced by (3·2) and (4·1) to consider nonconstant functions $\beta^\theta(s)$. Now turn to the conditional distribution of $Z(dx\,ds)$.

THEOREM 3. *Conditional on* θ, Γ *and* $N(dx)$, *the random measure* $Z(dx\,ds)$ *is distributed as the sum of* $N_{\mathscr{X}} \equiv N(\mathscr{X})$ *unit point masses at points* $(x_n, s_n) \in \mathscr{X} \times \mathscr{S}$, *where* x_n *ranges over the support of* $N(dx)$, *counted according to multiplicity, and where* $\{s_n\}$ *are the realised values of conditionally independent random variables* $\{S_n\}_{n \leqslant N_{\mathscr{X}}}$ *with discrete distributions*

$$\mathrm{pr}(S_n = \sigma_m | \theta, \Gamma, N) = v_m k^\theta(x_n, \sigma_m)/\Lambda(x_n)$$

for the sequence $\{(v_m, \sigma_m)\}$ *of* (2·2).

Proof. The representation $\Lambda(x) = \sum_{m<\infty} v_m k^\theta(x, \sigma_m)$ of the Poisson mean follows from (2·2) and (3·1), and the definitions of S_n and $Z(dx\,ds)$ complete the proof. □

The complete conditional distribution of θ, given Γ and $Z(dx\,ds)$, depends on the arbitrary prior distribution $\pi(d\theta)$ and so cannot in general lead to a simple sampling scheme; nevertheless it will have a density function, if $\pi(d\theta)$ does, that will be needed for a Metropolis step (Hastings, 1970) in the calculations to follow.

THEOREM 4. *Let* $\alpha^\theta(ds)$ *have a density function* $\alpha^\theta(s)$ *with respect to a fixed nonatomic probability measure* $\Pi(ds)$ *on* $\mathscr{S}$, *and* $\pi(d\theta)$ *a density function* $\pi(\theta)$ *with respect to some reference measure* $d\theta$ *on* Θ. *Then for* $M > N_{\mathscr{X}} \equiv N(\mathscr{X})$ *the conditional density of* θ, *given* $Z(dx\,ds)$ *and the truncated gamma measure* $\Gamma_M(ds)$ *constructed in Corollary* 1, *is proportional*

to

$$\theta|_{Z,\Gamma_M} \propto \pi(\theta) \exp\left[- \int_{\mathscr{S}} \log\left\{1 + \frac{k^\theta(\mathscr{X}, s)}{\beta^\theta(s)}\right\} \alpha^\theta(ds) - E_1\{v_M \beta^\theta_*(\sigma_M)\}\alpha^\theta(\sigma_M)\right]$$
$$\times \prod_{n \leqslant N_{\mathscr{X}}} k^\theta(x_n, s_n) \prod_{m \leqslant M} [\alpha^\theta(\sigma_m)^{i_m} \exp\{-v_m \beta^\theta_*(\sigma_m)\}],$$

where $\beta^\theta_*(s) \equiv \beta^\theta(s) + k^\theta(\mathscr{X}, s)$ *and* $i_m \equiv 1$ *if* $\sigma_m \notin \{s_n : n < m\}$, *otherwise* $i_m \equiv 0$.

Theorem 4 needs only a small adjustment to accommodate measures $\Pi(ds)$ with atoms: simply replace $\alpha^\theta(\sigma_m)^{i_m}$ in the final product by $\{\pi_m \alpha^\theta(\sigma_m) + \zeta_m\}$ for atoms σ_m with $\pi_m \equiv \Pi(\{\sigma_m\}) > 0$, where ζ_m is the cumulative multiplicity $\#\{n < m : s_n = \sigma_m\}$, but Remark 1 offers a more efficient way to accommodate atoms.

4·3. *The Markov chain Monte Carlo scheme*

Theorems 2–4 provide all the conditional distributions needed to implement a Markov chain Monte Carlo scheme for sampling from the joint posterior distribution of the uncertain quantities from the augmented model,

$$\theta \sim \pi(d\theta), \quad \Gamma(ds)|_\theta \sim \mathrm{Ga}\{\alpha^\theta(ds), \beta^\theta(s)^{-1}\}, \quad Z(dx\, ds)|_{\theta,\Gamma} \sim \mathrm{Po}\{k^\theta(dx, s)\Gamma(ds)\}.$$

Given a prior density function $\pi(\theta)$ on a parameter space Θ, a reference measure $w(dx)$ on the observation space $\mathscr{X}$, a probability measure $\Pi(ds)$ on an auxiliary space $\mathscr{S}$, a measurable kernel function $k^\theta(x, s)$, a gamma process shape density $\alpha^\theta(s)$ and inverse scale function $\beta^\theta(s)$ as in § 3·1, a Markov transition density $Q(\theta, \theta^*)$ on Θ, an integer $M \gg N \equiv N_{\mathscr{X}}$ and initial points $\theta^0 \in \Theta$ and $\{(v_m^0, \sigma_m^0)\}_{m \leqslant M} \subset \mathbb{R}_+ \times \mathscr{S}$, successive points can be generated by the following hybrid Gibbs/Metropolis scheme starting at iteration $t = 1$.

1. Gibbs step to update the augmentation points. Given θ^{t-1} and Γ^{t-1},
   (a) generate independent $S^t = \{s_n^t\}_{n \leqslant N}$ with $\mathrm{pr}(S_n^t = \sigma_m^{t-1}) \propto v_m^{t-1} k^{\theta^{t-1}}(x_n, \sigma_m^{t-1})$;
   (b) (optional) move S^t with a Metropolis/Hastings step; see note below.
2. Gibbs step to update the gamma process. Given θ^{t-1} and $S^t = \{s_n^t\}_{n \leqslant N}$,
   (a) set $\sigma_m^t \equiv s_m^t$ $(1 \leqslant m \leqslant N)$ and generate independent $\sigma_m^t \sim \Pi(ds)$ for $N < m \leqslant M$;
   (b) set $\alpha_m^t \equiv \alpha^{\theta^{t-1}}(\sigma_m^t)$ and $\beta_m^t \equiv \beta^{\theta^{t-1}}(\sigma_m^t) + k^{\theta^{t-1}}(\mathscr{X}, \sigma_m^t)$ for $1 \leqslant m \leqslant M$, and put $i_m^t \equiv 0$ if $s_n^t = \sigma_m^t$ for some $n < m$, otherwise $i_m^t \equiv 1$;
   (c) generate successive jumps $\{\tau_m\}_{m \leqslant M}$ of a standard Poisson process;
   (d) set

$$v_m^t \equiv \begin{cases} (\tau_m - \tau_{m-1})/\beta_m^t & \text{for } 1 \leqslant m \leqslant N, \text{ see Remark 1,} \\ E_1^{-1}\{(\tau_m - \tau_N)/\alpha_m^t\}/\beta_m^t & \text{for } N < m \leqslant M, \text{ see Corollary 1;} \end{cases}$$

   (e) set $\Gamma^t(ds) \equiv \sum_{m \leqslant M} v_m^t \delta_{\sigma_m^t}(ds)$.
3. Metropolis/Hastings step to update the parameter θ. Given θ^{t-1}, Γ^t and $S^t = \{s_n^t\}_{n \leqslant N}$,
   (a) set $\theta^- \equiv \theta^{t-1}$ and generate $\theta^+ \sim Q(\theta^-, \theta^+)$;
   (b) set $k_n^- \equiv k^{\theta^-}(x_n, s_n^t)$, $k_n^+ \equiv k^{\theta^+}(x_n, s_n^t)$;
   (c) set $\alpha_m^- \equiv \alpha^{\theta^-}(\sigma_m^t)$, $\alpha_m^+ \equiv \alpha^{\theta^+}(\sigma_m^t)$;
   (d) set $\beta_m^- \equiv \beta^{\theta^-}(\sigma_m^t) + k^{\theta^-}(\mathscr{X}, \sigma_m^t)$, $\beta_m^+ \equiv \beta^{\theta^+}(\sigma_m^t) + k^{\theta^+}(\mathscr{X}, \sigma_m^t)$;

(e) set

$$P^t \equiv \left\{\frac{\pi(\theta^+)}{\pi(\theta^-)}\right\}\left\{\frac{Q(\theta^+,\theta^-)}{Q(\theta^-,\theta^+)}\right\}\left\{\prod_{n\le N}\left(\frac{k_n^+}{k_n^-}\right)\right\}\exp\left[\sum_{m\le M}\left\{i_m^t\log\left(\frac{\alpha_m^+}{\alpha_m^-}\right)-v_m^t(\beta_m^+-\beta_m^-)\right\}\right]$$
$$\times\exp\left[-\int_{\mathscr{S}}\log\left\{1+\frac{k^{\theta^+}(\mathscr{X},s)}{\beta^{\theta^+}(s)}\right\}\alpha^{\theta^+}(ds)-E_1(v_M\beta_M^+)\alpha_M^+\right.$$
$$\left.+\int_{\mathscr{S}}\log\left\{1+\frac{k^{\theta^-}(\mathscr{X},s)}{\beta^{\theta^-}(s)}\right\}\alpha^{\theta^-}(ds)+E_1(v_M\beta_M^-)\alpha_M^-\right];$$

(f) set $\theta^t\equiv\theta^+$ with probability $(1\wedge P^t)$, otherwise set $\theta^t\equiv\theta^-=\theta^{t-1}$.

4. Set $t\leftarrow t+1$ and return to step 1.

The approach of this Markov chain to its equilibrium distribution can be slow, with some of the same augmentation points s_n^t recurring for long periods; convergence can be accelerated by adding another Metropolis step within Gibbs step 1(b), in which a small move is proposed for each s_n^t and is accepted or not following the Hastings rule (Tierney, 1994, p. 1704). A similar phenomenon affecting sampling from Dirichlet process distributions was recognised and overcome by Bush & MacEachern (1996) and MacEachern (1994).

5. EXAMPLES

5·1. *Description of the data*

As an illustration we analyse the density and spatial correlation of hickory trees in the Bormann research plot discussed in § 3·2, seeking answers to a number of questions. Is the tree distribution consistent with a homogeneous Poisson distribution? If the intensity is not constant, are intensities at nearby points positively associated? How might 'nearby' be determined?

The underlying dataset consists of a complete census count of every tree in the Bormann plot with diameter at height 1·5 m of at least 2·5 cm, including its diameter, to within 0·1 cm, its spatial coordinates, to within 0·1 m, and an indicator of its condition. The data, concerning 7859 trees of 38 species, were initially collected in 1951–1952 by Bormann and updated in 1974, 1982, 1989 and 1993 by Christensen, Peet and members of the Duke University and University of North Carolina botany departments in a continuing study of forest maturation (Bormann, 1953; Christensen, 1977). We group the four hickory species present and base our inference on the locations, shown as dots in Fig. 1(a), of the 85 hickory trees located at least 10 m from the boundary of the 140 m × 140 m plot with diameter at least 2·5 cm in the 1974 census, making no further use of the trees' species, diameters or conditions; see § 6 for possible extensions.

5·2. *Example: The continuous case*

We continue now with the model introduced as an illustration in § 3·2 with independent prior distributions, normal for θ_1 and θ_2 and log-normal for θ_3, chosen to give a prior mean density of about 60 trees per hectare, which is typical for this area, prior expected over-

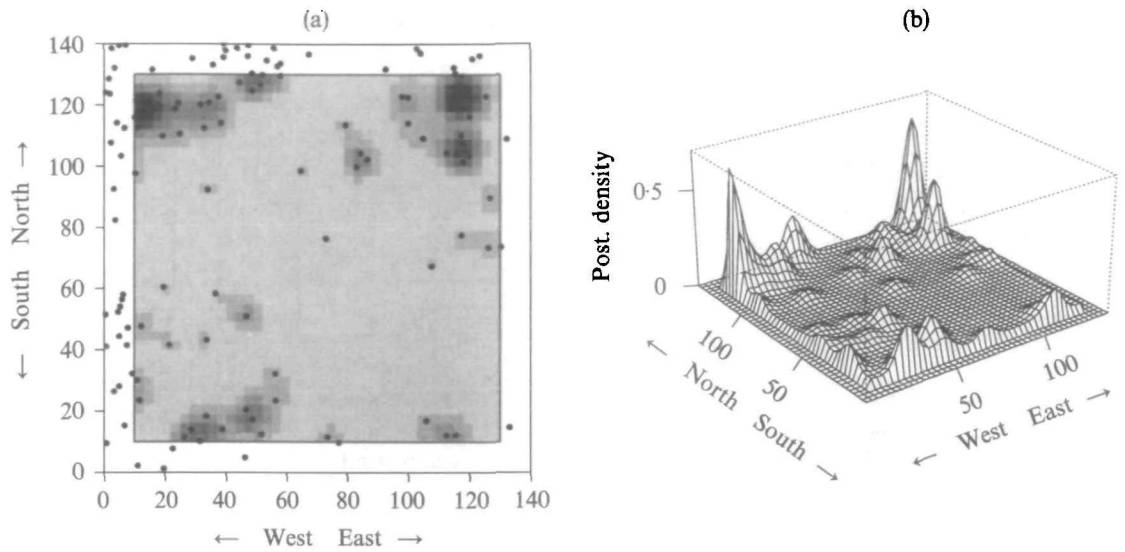

Fig. 1. Hickory tree data: Location on 140 metre × 140 metre plot. Image (a) and perspective (b) plots of posterior mean tree density $\Lambda(x)$.

dispersion of 50% and prior expected interaction distance of 5 m. Thus for $\theta \in \Theta \equiv \mathbb{R}^2 \times \mathbb{R}_+$,

$$\pi(d\theta) = \frac{1}{\theta_3(2\pi)^{3/2}} \exp\{-(\theta_1 + 4{\cdot}4)^2/2 - (\theta_2 - 8{\cdot}5)^2/2 - \log(\theta_3/5{\cdot}0)^2/2\}\, d\theta_1\, d\theta_2\, d\theta_3,$$

$$\alpha^\theta(ds) = e^{\theta_1}\, ds, \quad \beta^\theta(s)^{-1} \equiv e^{\theta_2}, \quad k^\theta(x, s) = \frac{1}{\pi\theta_3^2} \exp(-|x - s|^2/\theta_3^2), \quad m(dx) = 10^{-4}\, dx.$$

Figure 1 shows the posterior mean Poisson intensity $E\{\Lambda(x)\}$, as an image plot in Fig. 1(a) with darker regions indicating higher intensity and dots indicating the trees, and as a perspective plot in Fig. 1(b). Twenty-five thousand steps of the computational scheme of § 4·3 appear more than adequate for convergence, since time series plots of all the features we monitored had stabilised by 5000 steps; we used symmetric Gaussian random walks for both Metropolis steps 3(a) and 1(b), with step sizes chosen to ensure about 40% acceptance of the proposed steps.

Figure 2 shows the prior density, as a curve, and estimated posterior density, as a histogram, for the overall tree density $e^{\theta_1 + \theta_2}$, in trees per hectare, the overdispersion $10^{-4}e^{\theta_2}$, as a percentage, and the interaction distance θ_3, in metres, with the posterior means 76·2 trees/ha, 178·5% and 4·6 m indicated by vertical lines. This analysis suggests that there is substantially more overdispersion than anticipated and spatial interaction with positive association in tree density at distances of up to about 5 m.

5·3. *Example: The discrete case*

When both $\mathcal{X} = \{x_i\}_{i \in I}$ and $\mathcal{S} = \{s_j\}_{j \in J}$ are finite or countable, the models can be expressed in terms of vectors and matrices $N_i \equiv N(\{x_i\})$, $w_i \equiv w(\{x_i\})$, $\Lambda_i \equiv \Lambda(x_i)$, $\alpha_j^\theta \equiv \alpha^\theta(\{s_j\})$, $\beta_j^\theta \equiv \beta^\theta(s_j)$, $\Gamma_j \equiv \Gamma(\{s_j\})$, $Z_{ij} \equiv Z(\{x_i\} \times \{s_j\})$ and $k_{ij}^\theta \equiv k^\theta(x_i, s_j)$ in the simpler form

$$\theta \sim \pi(d\theta), \quad \{\Gamma_j\}_{j \in J}|_\theta \sim \mathrm{Ga}\{\alpha_j^\theta, (\beta_j^\theta)^{-1}\}, \quad \Lambda_i \equiv \sum_{j \in J} k_{ij}^\theta \Gamma_j, \quad \{N_i\}_{i \in I}|_{\theta,\Gamma} \sim \mathrm{Po}(\Lambda_i w_i).$$

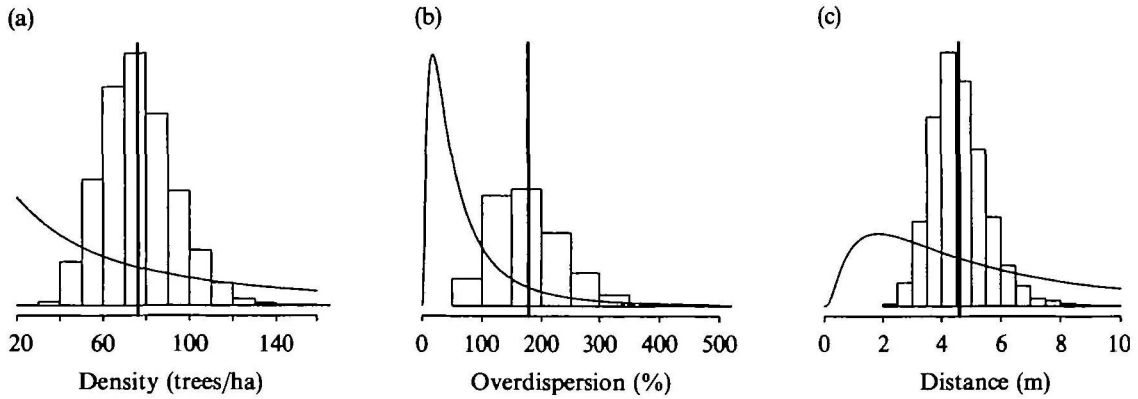

Fig. 2. Hickory tree data. Prior density functions and posterior histograms for (a) overall density (trees/hectare), (b) overdispersion (%), and (c) interaction distance (metres). Vertical lines indicate posterior means.

Again data augmentation, now from independent multinominal distributions

$$\{Z_{i\cdot}\}_{i\in I}|_{\theta,\Gamma,N} \sim \text{MN}(N_i, p^\theta_{i\cdot}), \quad p^\theta_{ij} \equiv k^\theta_{ij}\Gamma_j/\Lambda_i,$$

makes the model conjugate, but now sampling from the gamma distribution is routine and does not require the Inverse Lévy Measure algorithm to sample from a gamma random field as in Theorem 1; see Remark 1.

We now study the same data, but this time we divide the 120 m × 120 m area into a 6 × 6 lattice of 20 m × 20 m quadrats and aggregate the counts N_i of hickory trees in each quadrat. The empirical density plot Fig. 3(a) indicates these counts numerically and through shading.

Let I be a set indexing the 36 quadrats $\{x_i\}$ and view $\mathcal{X} = \{x_i\}_{i\in I}$ as a nearest neighbour adjacency graph; set $J_1 \equiv I$ and let J_2 be the set of unordered pairs (i, i') indexing the 60

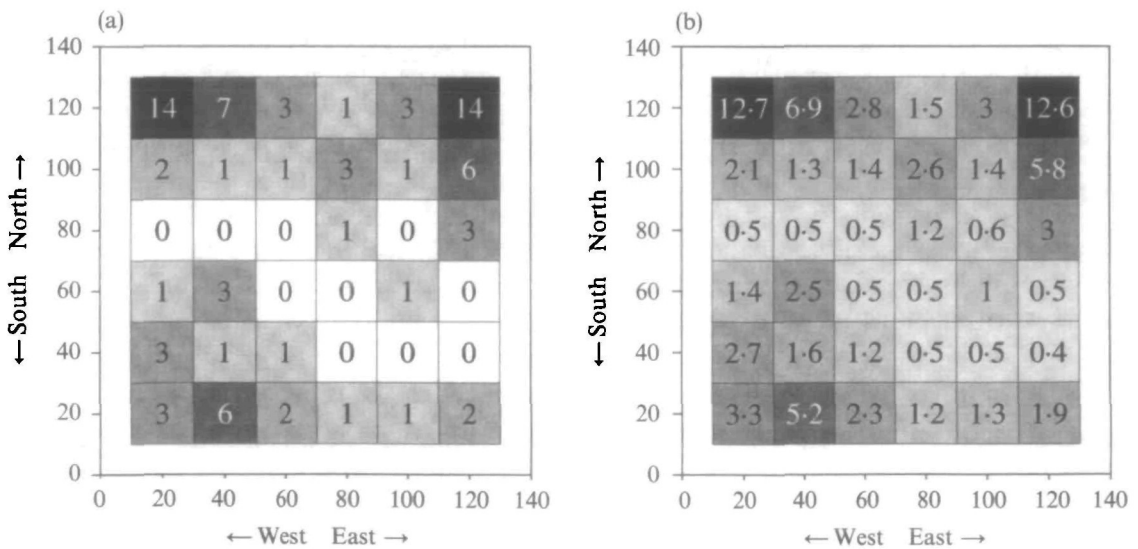

Fig. 3. Hickory tree data. Empirical (a) and posterior (b) image plots of intensity $\Lambda(x)$ on 6 × 6 quadrats in 120 metre × 120 metre area.

edges or nearest neighbour pairs $\{x_i, x_{i'}\}$ that comprise the dual graph $\mathscr{X}'$; and let Γ be a 96-dimensional gamma-distributed random vector indexed by $J \equiv (J_1 \cup J_2)$. We model the Poisson intensity Λ_i at x_i as the sum of a part associated with only that quadrat, $k^\theta_{ii}\Gamma_i = \theta_1\Gamma_i$, and up to four edge-terms $k^\theta_{ij}\Gamma_j = \theta_2\Gamma_j$, $j = (i, i') \in J_2$, shared with nearest neighbours. Figure 3(b) shows the estimated posterior mean intensity $E(\Lambda_i)$ for each quadrat after 10 000 Markov chain Monte Carlo steps, both numerically and by shading, based on independent standard log-normal prior distributions for θ_1 and θ_2. The spatial pattern apparent in the empirical density plot Fig. 3(a) remains, but is smoothed considerably in the posterior plot Fig. 3(b). Related discrete models with more conventional nearest neighbour and distance-based kernels are studied in more detail at a variety of resolutions in Ickstadt & Wolpert (1997).

6. Discussion

We now return to the questions posed at the beginning of § 5·1. In the Example of § 5·2 we see substantially more overdispersion than anticipated, in that the posterior probability of at least 100% overdispersion is 0·929, while the log-normal prior accords only probability 0·239 to the event; evidently the hickory trees are not spread uniformly throughout $\mathscr{X}$. The interaction distance θ_3 has posterior mean 4·59 m, with 90% of the mass in the range $3{\cdot}24 \text{ m} < \theta_3 < 6{\cdot}23$ m, consonant with the prior distribution, with median 5·0 m, but much more concentrated in that the prior probability of this interval is only 0·255; in particular, the posterior probability of interaction at distances exceeding 3 m (respectively 2 m) is 0·976 (respectively $> 0{\cdot}999$), offering overwhelming evidence of spatial interaction.

The models can be extended in a variety of ways. The kernel $k^\theta(x, s)$ and hence the Poisson intensity $\Lambda(x)$ may be modelled as uncertain functions of observed collateral information $C(x)$, such as location, soil type, elevation or aspect at sites $x \in \mathscr{X}$, to avoid mistaking covariate dependence or a spatial trend for spatial dependency; one convenient choice is the log-linear kernel $k^\theta(x, s) \equiv e^{C(x)\theta_C} k^\theta_0(x, s)$, with baseline kernel $k^\theta_0(x, s)$. Individual-specific covariates, such as the species, diameter and condition of each tree, may be included by treating the observations as a marked point process, a random measure on the Cartesian product $(\mathscr{X} \times \mathscr{A})$ of the observation space $\mathscr{X}$ with some space $\mathscr{A}$ of possible attributes, or 'marks', so that the intensities and relative abundances for the several tree species or hickory subspecies can be considered simultaneously in a spatial study of biodiversity. Temporal trends may be studied by treating the parameter $\theta = \theta(t) \in \Theta$ as a time series or stochastic process, to permit the study of evolving spatial trends; the surveys from 1952, 1974, 1982, 1989 and 1993 might be used together to make inference about the time-dependence of $\theta(t)$ and so reveal trends and support predictions. Some of these possibilities are pursued by Ickstadt & Wolpert (1997) in a discrete setting.

The latent random field $\Gamma(ds)$ may itself be of independent interest. A mechanistic model, in which Γ represents the unobserved distribution of soil nutrients which diffuse over time to sites $x \in \mathscr{X}$, would be quite similar to the present model, in that the Gaussian kernel $k^\theta(x, s)$ is also the Fickian diffusion kernel (Crank, 1975, p. 28). This suggests that a more elaborate analysis might include eliciting features of nonconstant densities $\alpha^\theta(s)$ and scales $\beta^\theta(s)^{-1}$, reflecting the anticipated distribution of nutrients, minerals, water and other growth requirements.

The doubly stochastic Bayesian hierarchical models introduced here are applicable in a wide range of problems, both continuous and discrete, in any number of dimensions, that feature correlated count data: disease mapping, where the counts are cases, bioabund-

ance, where they are individuals, network analysis, where they are packets passing a node, and survival analysis, where they are failure times and $\mathcal{X} = \mathbb{R}_+$ represents time. In ecological and epidemiological applications it is common to study count data aggregated within political units while environmental covariates are reported at different levels of aggregation; the existence of a continuous random field model underlying the discrete version of the present models enables one to accommodate data and covariates reported at varying, and sometimes incommensurate, spatial levels without any further aggregation.

The Inverse Lévy Measure algorithm and sampling method of Theorem 1, with its completely arbitrary location-sampling distribution $\Pi(ds)$ and its easy extensibility to arbitrary spatial Lévy processes, and the hybrid Metropolis/Gibbs approach, with a completely arbitrary prior distribution $\pi(d\theta)$, free the modeller from the need to choose conventional or merely convenient prior distributions and intensity measures. Software and data are available on request from the authors.

ACKNOWLEDGEMENT

The authors would like to thank Norm Christensen and Robert Peet for sharing data and insight, Jürgen Lehn for support and encouragement and Merlise Clyde, Noel Cressie, Michael Lavine and Philip Protter for helpful conversations. We are grateful to the Editor, Associate Editor and two referees for their careful reading and valuable suggestions. This work was supported by the United States Environmental Protection Agency, United States National Science Foundation and the Deutsche Forschungsgemeinschaft and was performed while the second author was visiting Duke University.

APPENDIX

Proofs

Proof of Theorem 1. First consider simple functions $\phi(s)$, $\alpha(s)$ and $\beta(s)$, all constant on elements of a fixed partition $\{B_j\}_{j\in J}$ of $\mathcal{S}$ with values $\phi_j \geqslant 0$, $\alpha_j \geqslant 0$ and $\beta_j > 0$, respectively. For each $j \in J$ set $p_j \equiv \Pi(B_j)$ and $P_t^{(j)} \equiv \sum_{m:p_j\tau_m \leqslant t} 1_{B_j}(\sigma_m)$. Then $\{P_t^{(j)}\}_{j\in J}$ are independent standard Poisson processes, each a time-changed Bernoulli-thinned version of $P_t \equiv \sum_{m:\tau_m\leqslant t} 1$, with jump times τ_m^j, and

$$\Gamma(B_j) = \sum_{m:\sigma_m \in B_j} \left\{E_1^{-1}\left(\frac{\tau_m}{\alpha_j}\right)\beta_j^{-1}\right\} = \sum_{m<\infty} \left\{E_1^{-1}\left(\frac{\tau_m^j}{p_j\alpha_j}\right)\beta_j^{-1}\right\}$$

are independent gamma-distributed $\mathrm{Ga}(p_j\alpha_j, \beta_j^{-1})$ random variables, so

$$-\log E(e^{-\Gamma[\phi]}) = -\log \prod_{j\in J}\left(1+\frac{\phi_j}{\beta_j}\right)^{-p_j\alpha_j} = \int_{\mathcal{S}} \log\left\{1+\frac{\phi(s)}{\beta(s)}\right\}\alpha(ds)$$

is the Laplace exponent for the gamma process. A limiting argument completes the proof for any nonnegative measurable $\phi(s)$ integrable with respect to $\beta(s)^{-1}\alpha(ds)$, while $\Gamma[\phi] = \infty$ almost surely, and $E(e^{-\Gamma[\phi]}) = 0$, for any ϕ not so integrable. □

Proof of Corollary 1. For $\phi(s) \geqslant 0$ the conditional distribution, given Γ_M, of the necessarily positive truncation error $(\Gamma[\phi] - \Gamma_M[\phi]) = \sum_{m>M} v_m\phi(\sigma_m)$ depends only on τ_M, with conditional expectation

$$\begin{aligned} E(\Gamma[\phi] - \Gamma_M[\phi] \mid \Gamma_M) &= \int_{\mathcal{S}} \phi(s)\left(1 - \exp\left[-E_1^{-1}\left\{\frac{\tau_M}{\alpha(s)}\right\}\right]\right)\beta(s)^{-1}\alpha(ds) \\ &\leqslant \int_{\mathcal{S}} \phi(s)E_1^{-1}\left\{\frac{\tau_M}{\alpha(s)}\right\}\beta(s)^{-1}\alpha(ds) \leqslant \int_{\mathcal{S}} \phi(s)(e^{\tau_M/\alpha(s)}-1)^{-1}\beta(s)^{-1}\alpha(ds), \end{aligned}$$

where the last bound follows from the inequality $e^x E_1(x) \leqslant \log(1 + 1/x)$ of Abramowitz & Stegun (1964, § 5.1.20). □

Proof of Theorem 2. Conditional on θ and Γ, $Z(dx\,ds)$ is a Poisson measure on $\mathscr{X} \times \mathscr{S}$ with intensity measure $k^\theta(dx, s)\Gamma(ds)$; thus $(X_n, S_n)|_{\theta,\Gamma} \sim k^\theta(dx_n, s_n)\Gamma(ds_n)/\Lambda(\mathscr{X})$. First consider simple functions $\beta^\theta(s)$ and $k^\theta(dx, s)$, constant on a fixed partition $\{B_j\}_{j\in J}$ of $\mathscr{S}$; fix any $s_j \in B_j$, and set $\alpha_j^\theta \equiv \alpha^\theta(B_j)$, $\beta_j^\theta \equiv \beta^\theta(s_j)$, $\Gamma_j \equiv \Gamma(B_j)$, with realisations γ_j, and $z_j = Z_2(B_j)$. Let $\phi(s)$ be nonnegative and simple as well, with constant value ϕ_j on each B_j; note that $\Lambda(\mathscr{X}) = \sum_{j\in J} k^\theta(\mathscr{X}, s_j)\Gamma_j$ and $\Gamma[\phi] = \sum_{j\in J}\phi_j\Gamma_j$. The joint distribution is then proportional to

$$\theta, \Gamma, Z \sim \pi(d\theta)\left\{\frac{\Lambda(\mathscr{X})^{N_{\mathscr{X}}} e^{-\sum_{j\in J} k^\theta(\mathscr{X}, s_j)\gamma_j}}{N_{\mathscr{X}}!}\right\}\prod_{j\in J}\left\{\frac{(\beta_j^\theta)^{\alpha_j^\theta}\gamma^{\alpha_j^\theta - 1}e^{-\beta_j^\theta\gamma_j}}{\Gamma(\alpha_j^\theta)}\,d\gamma_j\right\}$$
$$\times \prod_{n\leqslant N_{\mathscr{X}}}\left\{\frac{k^\theta(x_n, s_n)w(dx_n)\sum_{j\in J} 1_{B_j}(s_n)\gamma_j\Gamma(ds_n | s_n \in B_j)}{\Lambda(\mathscr{X})}\right\}, \tag{A·1}$$

so that

$$E^\theta(e^{-\Gamma[\phi]}|\theta, Z) \propto \prod_{j\in J}\left(\int_0^\infty e^{-\phi_j\gamma_j}\gamma^{\alpha_j^\theta - 1}e^{-\beta_j^\theta\gamma_j}e^{-k^\theta(\mathscr{X}, s_j)\gamma_j}\gamma_j^{z_j}\,d\gamma_j\right) \propto \prod_{j\in J}\{\phi_j + \beta_j^\theta + k^\theta(\mathscr{X}, s_j)\}^{-\alpha_j^\theta - z_j}.$$

The proportionality constant is determined by the relation $e^{-\Gamma[0]} \equiv 1$, giving

$$-\log E^\theta(e^{-\Gamma[\phi]}|\theta, Z) = -\log\prod_{j\in J}\left\{\frac{\phi_j + \beta_j^\theta + k^\theta(\mathscr{X}, s_j)}{\beta_j^\theta + k^\theta(\mathscr{X}, s_j)}\right\}^{-\alpha_j^\theta - z_j}$$
$$= \int_{\mathscr{S}} \log[1 + \phi(s)\{\beta^\theta(s) + k^\theta(\mathscr{X}, s)\}^{-1}](\alpha^\theta + Z_2)(ds),$$

the Laplace exponent given in (2·1) for the gamma process claimed in (4·1). A routine passage to the limit, refining the partition B_j, completes the proof. □

Proof of Theorem 4. Integrate (A·1) with respect to $\{\gamma_j\}_{j\in J}$ to find

$$\theta, Z \sim \frac{\pi(\theta)}{N_{\mathscr{X}}!}\prod_{j\in J}\left[\frac{(\beta_j^\theta)^{\alpha_j^\theta}}{\{\beta_j^\theta + k^\theta(\mathscr{X}, s_j)\}^{\alpha_j^\theta + z_j}}\right]\prod_{j\in J}\left\{\frac{\Gamma(\alpha_j^\theta + z_j)}{\Gamma(\alpha_j^\theta)}\right\}\prod_{n\leqslant N_{\mathscr{X}}}\{k^\theta(x_n, s_n)w(dx_n)\}.$$

As the partition $\{B_j\}$ is refined, the first product converges to

$$\exp\left[-\int_{\mathscr{S}}\log\left\{1 + \frac{k^\theta(\mathscr{X}, s)}{\beta^\theta(s)}\right\}\alpha^\theta(ds)\right]\prod_{n\leqslant N_{\mathscr{X}}}\{\beta^\theta(s_n) + k^\theta(\mathscr{X}, s_n)\}^{-1}$$

and the second, by nonatomicity, to const $\times \prod_{n\in N_0}\{\alpha^\theta(s_n)\Pi(ds_n)\}$, where the product extends only over a set N_0 indexing the distinct augmentation points s_n. After terms are combined, the joint density with respect to the product measure $d\theta\prod_{n\leqslant N_{\mathscr{X}}} w(dx_n)\prod_{n\in N_0}\Pi(ds_n)$ is proportional to

$$\theta, Z \propto \pi(\theta)\exp\left[-\int_{\mathscr{S}}\log\left\{1 + \frac{k^\theta(\mathscr{X}, s)}{\beta^\theta(s)}\right\}\alpha^\theta(ds)\right]\prod_{n\leqslant N_{\mathscr{X}}}\left\{\frac{k^\theta(x_n, s_n)}{\beta^\theta(s_n) + k^\theta(\mathscr{X}, s_n)}\right\}\prod_{n\in N_0}\alpha^\theta(s_n).$$

By Theorem 2 and Remark 1 the conditional distribution of $\Gamma(ds)$ may be written in form (2·2) with $\sigma_m = s_m$ and $v_m \equiv (\tau_m - \tau_{m-1})\beta_*^\theta(\sigma_m)^{-1}$, for $m \leqslant N_{\mathscr{X}}$, $\sigma_m \sim \Pi(ds)$ and

$$v_m \equiv E_1^{-1}\{(\tau_m - \tau_{N_{\mathscr{X}}})/\alpha^\theta(\sigma_m)\}\beta_*^\theta(\sigma_m)^{-1},$$

for $m > N_{\mathscr{X}}$, for jumps τ_m of a standard unit-rate Poisson process; by Corollary 2, the θ-dependent

portion of the density function of the first M points (v_m, σ_m) is proportional to

$$\Gamma_M|_{\theta,Z} \propto \prod_{1 \le n \le N_{\mathcal{Z}}} \beta^{\theta}_*(s_n) \prod_{N_{\mathcal{Z}} < m \le M} \alpha^{\theta}(\sigma_m) \exp\left[- \sum_{1 \le m \le M} v_m \beta^{\theta}_*(\sigma_m) - E_1\{v_M \beta^{\theta}_*(\sigma_M)\} \alpha^{\theta}(\sigma_M) \right].$$

Combining the last two equations completes the proof. □

REFERENCES

ABRAMOWITZ, M. & STEGUN, I. A. (Ed.). (1964). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, Applied Mathematics Series, **55**. Washington, DC: National Bureau of Standards.

ANTONIAK, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2**, 1152–74.

BADDELEY, A. J. & VAN LIESHOUT, M. N. M. (1995). Area-interaction point processes. *Ann. Inst. Statist. Math.* **47**, 601–19.

BERNARDINELLI, L., PASCUTTO, C., BEST, N. G. & GILKS, W. R. (1997). Disease mapping with errors in covariates. *Statist. Med.* **16**, 741–52.

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with Discussion). *J. R. Statist. Soc.* B **35**, 192–236.

BESAG, J., YORK, J. & MOLLIÉ, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with Discussion). *Ann. Inst. Statist. Math.* **43**, 1–59.

BONDESSON, L. (1982). On simulation from infinitely divisible distributions. *Adv. Appl. Prob.* **14**, 855–69.

BORMANN, F. H. (1953). The statistical efficiency of sample plot size and shape in forest ecology. *Ecology* **34**, 474–87.

BUSH, C. A. & MACEACHERN, S. N. (1996). A semi-parametric Bayesian model for randomised block designs. *Biometrika* **83**, 275–85.

CHRISTENSEN, N. L. (1977). Changes in structure, pattern and diversity associated with climax forest maturation in Piedmont, North Carolina. *Am. Midland Naturalist* **97**, 176–88.

CLAYTON, D. G. & KALDOR, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671–81.

COX, D. R. (1955). Some statistical methods connected with series of events (with Discussion). *J. R. Statist. Soc.* B **17**, 129–64.

COX, D. R. & ISHAM, V. (1980). *Point Processes*. New York: Chapman and Hall.

CRANK, J. (1975). *The Mathematics of Diffusion*, 2nd ed. Oxford: Oxford University Press.

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. New York: Wiley.

CRESSIE, N. A. C. & CHAN, N. H. (1989). Spatial modeling of regional variables. *J. Am. Statist. Assoc.* **84**, 393–401.

DAMIEN, P., LAUD, P. W. & SMITH, A. F. M. (1995). Approximate random variate generation from infinitely divisible distributions with applications to Bayesian inference. *J. R. Statist. Soc.* B **57**, 547–63.

DIGGLE, P. J. (1983). *Statistical Analysis of Spatial Point Patterns*. New York: Academic Press.

DYKSTRA, R. L. & LAUD, P. W. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9**, 356–67.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–30.

FERGUSON, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, 615–29.

FERGUSON, T. S. & KLASS, M. J. (1972). A representation of independent increment processes without Gaussian components. *Ann. Math. Statist.* **43**, 1634–43.

GELFAND, A. E. & SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Statist. Assoc.* **85**, 398–409.

GILKS, W. R., RICHARDSON, S. & SPIEGELHALTER, D. J. (1996). *Markov Chain Monte Carlo in Practice*. New York: Chapman and Hall.

HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

ICKSTADT, K. & WOLPERT, R. L. (1997). Multiresolution assessment of forest inhomogeneity. In *Case Studies in Bayesian Statistics, Volume 3*, Lecture Notes in Statistics, **121**, Ed. C. Gatsonis, J. S. Hodges, R. E. Kass, R. McCulloch, P. Rossi and N. D. Singpurwalla, pp. 371–86. New York: Springer-Verlag.

JACOD, J. & SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes*, Grundlehren der mathematischen Wissenschaften, **288**. Berlin: Springer-Verlag.

KARR, A. F. (1991). *Point Processes and their Statistical Inference*, 2nd ed., Probability, Pure and Applied, **7**. New York: Marcel Dekker.

KELLY, F. P. & RIPLEY, B. D. (1976). A note on Strauss's model for clustering. *Biometrika* **63**, 357–60.

LAUD, P. W., SMITH, A. F. M. & DAMIEN, P. (1996). Monte Carlo methods for approximating a posterior hazard rate process. *Statist. Comp.* **6**, 77–83.

LÉVY, P. (1937). *Théorie de L'Addition des Variables Aléatoires.* Paris: Gauthier-Villars.

MACEACHERN, S. N. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Commun. Statist.* B **23**, 727–41.

MØLLER, J., SYVERSVEEN, A. R. & WAAGEPETERSEN, R. P. (1998). Log Gaussian Cox processes. *Scand. J. Statist.* **25**. To appear.

NEYMAN, J. & SCOTT, E. L. (1958). Statistical approach to problems of cosmology (with Discussion). *J. R. Statist. Soc.* B **20**, 1–43.

RIPLEY, B. D. (1977). Modelling spatial patterns (with Discussion). *J. R. Statist. Soc.* B **39**, 172–212.

RIPLEY, B. D. & KELLY, F. P. (1977). Markov point processes *J. Lond. Math. Soc.* **15**, 188–92.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability. New York: Chapman and Hall.

STRAUSS, D. J. (1975). A model for clustering. *Biometrika* **62**, 467–76.

TANNER, M. A. & WONG, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Am. Statist. Assoc.* **82**, 528–50.

TIERNEY, L. (1994). Markov chains for exploring posterior distributions (with Discussion). *Ann. Statist.* **22**, 1701–62.

[*Received January* 1996. *Revised September* 1997]