BIOMETRICS 55, 738–745
September 1999

Modeling a Poisson Forest in Variable Elevations: A Nonparametric Bayesian Approach

Juha Heikkinen
Finnish Forest Research Institute, Unioninkatu 40 A, FIN-00170 Helsinki, Finland
email: Juha.Heikkinen@metla.fi

and

Elja Arjas
Rolf Nevanlinna Institute, P.O. Box 4, FIN-00014 University of Helsinki, Finland
email: Elja.Arjas@rni.helsinki.fi

SUMMARY. A nonparametric Bayesian formulation is given to the problem of modeling nonhomogeneous spatial point patterns influenced by concomitant variables. Only incomplete information on the concomitant variables is assumed, consisting of a relatively small number of point measurements. Residual variation, caused by other unmeasured influential factors, is modeled in terms of a spatially varying baseline intensity function. A Markov chain Monte Carlo scheme is proposed for the simultaneous nonparametric estimation of each unknown function in the model. The suggested method is illustrated by reanalysing a data set in Rathbun (1996, *Biometrics* **52,** 226–242), and the estimated models are compared with those obtained by Rathbun.

KEY WORDS: Ecological response curves; Nonparametric Bayesian inference; Reversible jump MCMC; Spatial interpolation; Spatial point process.

1. Introduction

In plant ecology, it is a natural idea to relate the spatial variation in abundance of plants to the values of one or more locally measured concomitant variables, such as ground elevation or acidity of soil, e.g. Information on such concomitant variables is, in practice, often incomplete, being restricted to a relatively small number of point measurements. Furthermore, other unmeasured influential factors are likely to be present, giving rise to greater variability in abundance than could be expected purely on grounds of the measured concomitant variables. Such residual variation is expressed naturally in terms of a spatially varying baseline intensity function. Both the dependence on measured concomitant variables and the baseline are unlikely to have a known functional form, however. Such considerations lead to a nonparametric model formulation, where one faces the additional task of estimating the true values of the considered concomitant variable(s). Here this problem is given a nonparametric Bayesian formulation, and a Markov chain Monte Carlo (MCMC) scheme is proposed for the numerical estimation.

As an illustration of the method, we reanalyse a data set described in Rathbun (1996). It contains the mapped locations of trees of several species in a 250-m × 200-m region of Titi Hammock, a beech–magnolia forest in southern Georgia, USA. The stand patterns show clear dependence on the ground elevation, of which measurements were taken on a regular square lattice of 11 × 9 sample points 25 m apart. Consider, e.g., the locations of individual ironwood *Carpinus caroliniana* trees shown in Figure 1. Comparison with the elevations (Table 1) reveals that ironwood is clearly most abundant in the lowest elevations.

Methodologically, this work stems from Arjas and Heikkinen (1997) and Heikkinen and Arjas (1998), where we developed a new method for the nonparametric Bayesian estimation of the intensity of a nonhomogeneous Poisson process. The method is based on mixing variable dimensional piecewise constant approximations, whose smoothness is regulated by a conditional autoregressive prior distribution. It was noted in the discussion of the latter paper that the same approach is applicable much more generally to the estimation of curves and surfaces and that these in turn can be parts of a larger, more realistic model. The current example gives a practical illustration of such analysis.

Our model has three unknown functions to be estimated. First, the ground elevation surface must be interpolated between the sample points; here we neglect the elevation measurement error mainly for the purpose of illustrating how interpolation can be tackled by our approach. Second, the response of the tree intensity to changes in elevation is an example of the usual regression curve. Finally, the baseline intensity surface is similar to the functions estimated in Heikkinen and Arjas (1998). Hence, this application provides a good oppor-

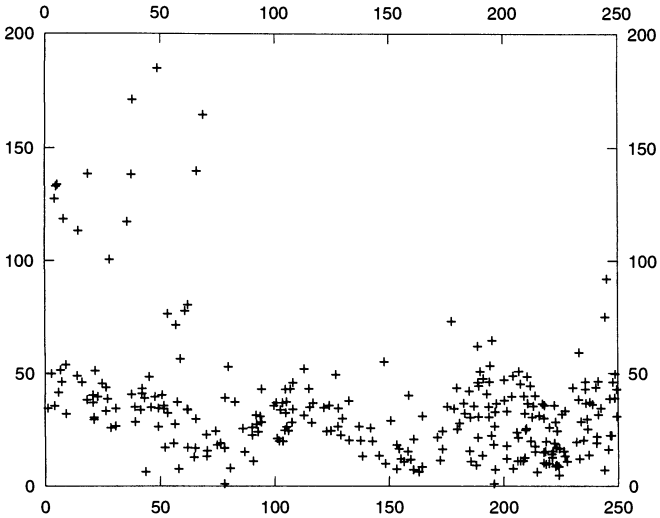

Figure 1. Locations of ironwood in the Titi Hammock study region; data kindly provided by Stephen Rathbun.

tunity to illustrate how basically a single estimation approach can be applied to a variety of problems, usually attacked by quite different frequentist methods.

Some minor changes are introduced to the method of Heikkinen and Arjas (1998). Most notably, the proper multinormal prior is replaced by an improper pairwise difference prior (e.g., Besag et al., 1995) and its precision parameter is treated as unknown.

The outline of this paper is as follows. In Section 2, we specify the statistical model to be considered. Some nonstandard aspects of the suggested MCMC estimation scheme are briefly discussed in Section 3, with more details given in the Appendix. In Section 4, we illustrate some of the estimation results and compare them with those of Rathbun (1996). Finally, we discuss the applied modeling and inferential ideas more generally in Section 5.

2. Model

In Rathbun (1996), the forest stand patterns were modeled using the modulated Poisson process, with the intensity taking a simple parametric form as a function of the elevation. The method proposed there for the maximum likelihood estimation was based on replacing the unobserved true elevations by their kriging predictors. Here we extend this model by allowing for a nonparametric response function and an additional unmeasured baseline intensity surface.

To introduce our model in more detail, let us denote the Titi Hammock study region $[0, 250] \times [0, 200]$ by E. Following Rathbun (1996), we model the observed tree pattern $\mathbf{x} = (x_1, \ldots, x_N) \subset E$ as a realisation of a nonhomogeneous Poisson process on E with an unknown intensity function $\lambda : E \to [0, \infty)$. Hence, the likelihood $p(\mathbf{x} \mid \lambda)$ of data $\mathbf{x}$ given an intensity function λ is proportional to

$$\exp\left\{-\int_E \lambda(s)\,\nu(ds)\right\} \prod_{n=1}^{N} \lambda(x_n), \tag{2.1}$$

where ν is the (two-dimensional) Lebesgue measure.

The intensity λ in turn is modeled as

$$\lambda(s) = \lambda_0(s)\gamma\{\zeta(s)\}, \tag{2.2}$$

where $\lambda_0 : E \to [0, \infty)$ is an unknown baseline intensity surface, $\zeta : E \to \mathbb{R}$ is the partially observed elevation surface, and $\gamma : \mathbb{R} \to [0, \infty)$ is an unknown multiplicative response of the intensity to changes in the elevation. The elevation measurements $\mathbf{z} = (z_1, \ldots, z_M)$ at the sample points $\mathbf{s} = (s_1, \ldots, s_M) \subset E$ are assumed to have negligible error so that $\zeta(s_m) = z_m$, or

$$p(\mathbf{z} \mid \zeta) \propto \prod_{m=1}^{M} \mathbf{1}\{z_m = \zeta(s_m)\}. \tag{2.3}$$

Following Arjas and Heikkinen (1997) and Heikkinen and Arjas (1998), the prior distribution of each unknown curve or surface f, $f \in \{\lambda_0, \zeta, \gamma\}$, is defined over a set of piecewise constant functions on a random partition of the function domain (E for surfaces λ_0 and ζ and $\mathbb{R}$ for curve γ). These partitions are obtained as Voronoi tessellations $\mathcal{E}(\boldsymbol{\xi}_f) = \{E_1(\boldsymbol{\xi}_f), \ldots, E_{K_f}(\boldsymbol{\xi}_f)\}$ of E or $\mathbb{R}$, generated by random point patterns $\boldsymbol{\xi}_f = (\xi_{f,1}, \ldots, \xi_{f,K_f})$. A realisation from the prior distribution is then parametrised by the generating point pattern $\boldsymbol{\xi}_f$ and the function values $\boldsymbol{\eta}_f = (\eta_{f,1}, \ldots, \eta_{f,K_f})$ in each Voronoi tile; the intensities are modeled in log-scale, whereby $\eta_{\lambda_0,k}$ denotes the value of $\log \lambda_0$ in $E_k(\boldsymbol{\xi}_{\lambda_0})$ and $\eta_{\gamma,k}$ the value of $\log \gamma$ in $E_k(\boldsymbol{\xi}_\gamma)$. Note that, due to the randomness of the

Table 1

Relative elevations (in meters) at sample points in Titi Hammock study area; published in Rathbun (1996)

	X										
Y	0	25	50	75	100	125	150	175	200	225	250
200	7.1	8.4	7.4	5.2	5.3	7.7	9.9	12.4	13.3	14.3	14.6
175	6.8	6.9	4.6	5.4	7.3	8.6	9.8	10.5	11.4	11.3	11.4
150	4.3	4.2	4.5	6.4	7.8	8.6	8.6	9.0	8.8	8.8	8.8
125	3.5	5.1	4.9	6.3	7.7	7.5	7.1	7.2	6.6	6.5	6.5
100	4.2	4.6	5.1	5.6	6.6	5.8	4.7	4.3	4.1	4.1	4.2
75	3.3	4.2	4.6	4.6	4.6	4.1	3.1	2.8	2.6	2.2	2.1
50	2.6	2.2	2.4	3.7	2.2	2.2	2.2	1.2	0.1	0.4	−0.4
25	2.3	1.7	1.5	1.2	0.8	0.8	0.9	0.5	0.4	0.0	−0.4
0	7.3	6.3	4.9	5.1	5.0	3.8	1.0	2.6	0.5	1.0	2.9

partitions, the pointwise posterior means do not need to form a piecewise constant function and, indeed, the posterior mean curves and surfaces are typically smooth continuous functions. (For further discussion on dynamic step function approximations in Bayesian inference, cf., Arjas and Andreev (1996) or Heikkinen (1998).)

The prior distribution of f is now determined by specifying the joint prior of $\boldsymbol{\xi}_f$ and $\boldsymbol{\eta}_f$. This is done via the chain rule decomposition $p(\boldsymbol{\xi}_f, \boldsymbol{\eta}_f) = p(\boldsymbol{\xi}_f)p(\boldsymbol{\eta}_f \mid \boldsymbol{\xi}_f)$. The prior of $\boldsymbol{\xi}_f$ is taken to be the homogeneous Poisson process with intensity $\lambda_{\boldsymbol{\xi},f}$, constrained to have at least two points in order for a pairwise difference prior to be sensibly defined for $\boldsymbol{\eta}_f$ given $\boldsymbol{\xi}_f$. Hence, the density $p(\boldsymbol{\xi}_f)$ is proportional to $\lambda_{\boldsymbol{\xi},f}^{K_f}\mathbf{1}(K_f > 1)$.

To build a smoothing prior for $\boldsymbol{\eta}_f$ given $\boldsymbol{\xi}_f$, we define a (realisation-dependent) neighbourhood relation $\underset{\boldsymbol{\xi}_f}{\sim}$ based on the partition $\mathcal{E}(\boldsymbol{\xi}_f)$. In one dimension, adjacent generating points are defined to be neighbours, and in two dimensions, two generating points are neighbours if the corresponding Voronoi tiles are contiguous, sharing a common edge. We then set up a Gaussian pairwise difference prior

$$p(\boldsymbol{\eta}_f \mid \boldsymbol{\xi}_f, \tau_f) \propto \left\{ \prod_{k=1}^{K_f} (\tau_f w_{k+}/2\pi)^{1/2} \right\} \times \exp\left\{ -\frac{\tau_f}{2} \sum_{k<j} w_{kj}(\eta_{f,k} - \eta_{f,j})^2 \right\}, \quad (2.4)$$

where the precision parameter τ_f determines the degree of smoothness, the weights $w_{kj} = w_{kj}(\boldsymbol{\xi}_f)$ are nonzero for neighbouring sites $\xi_{f,k} \underset{\boldsymbol{\xi}_f}{\sim} \xi_{f,j}$, and $w_{k+} = \Sigma_j w_{kj}$. This prior is improper because the density is invariant under shifts of every coordinate by the same amount, but the posterior is proper in the presence of any informative data. Also, the full conditionals

$$p(\eta_{f,k} \mid \boldsymbol{\eta}_{f,-k}, \boldsymbol{\xi}_f, \tau_f) \propto \exp\left\{ -\frac{\tau_f w_{k+}}{2} \left(\eta_{f,k} - \sum_j \frac{w_{kj}}{w_{k+}} \eta_{f,j} \right)^2 \right\} \quad (2.5)$$

are proper Gaussian distributions with the expected values given by weighted averages of the neighbouring levels.

The precision parameter τ_f is treated as unknown with an exponential prior $p(\tau_f) \propto \exp(-\beta_f \tau_f)$, and the generating point intensity $\lambda_{\boldsymbol{\xi},f}$ is used as a control variable to adjust the degree of smoothing. In the examples, we have applied distance-dependent weights $w_{kj}(\boldsymbol{\xi}) = |\xi_k - \xi_j|^{-1}$, both in one and two dimensions (some alternative schemes are discussed and applied in Arjas and Heikkinen (1997) and in Heikkinen and Arjas (1998)).

It should be noted that without any additional constraints the multiplicative intensity components λ_0 and γ are unidentifiable. The pairwise difference priors applied here have the convenient property (as compared to proper multinormal priors) that the posterior is invariant under transformations of the form $(\lambda_0 \to C\lambda_0, \gamma \to C^{-1}\gamma)$ for any positive constant C. Hence, realisations drawn from the posterior can be scaled afterwards in a way that best supports the desired interpretation.

In our notation, the model of Rathbun (1996) assumes λ_0 to be identically equal to one and $\log(\gamma)$ to have a polynomial form (both linear and quadratic functions are applied). To obtain comparable response function estimates, we can, e.g., scale the integrated baseline intensity per unit area in region E, given by $\int_E \lambda_0(s)\,\nu(ds)/\nu(E)$, to be identically equal to one in all realisations of λ_0. The intuition behind this choice is that, in the hypothetical case where the ground elevation in the entire region E were a constant z_0, the total number of trees in that region would be a Poisson random variable with mean $\gamma(z_0)\nu(E)$, i.e., $\gamma(z_0)$ would appear as a proportionality constant multiplying the area of E.

Another possibility for choosing the scaling is to decide on a reference level of elevation, say z_0, and set $\gamma(z_0) = 1$. Now, if the elevation had the constant value z_0 throughout the region E, $N(A)$ would be Poisson with parameter $\int_A \lambda_0(s)\,\nu(ds)$, and if this value were changed to the level z, the parameter would have to be multiplied by $\gamma(z)$. Thus, the interpretation of γ would be similar to that of the relative risk function in the proportional hazards model for survival.

3. Posterior Sampling

Our inferences are based on reversible jump Markov chain Monte Carlo sampling (Green, 1995) from the posterior distribution. In each basic update step, either a new value for a randomly selected $\eta_{f,k}$ or τ_f, the birth of a new marked generating point (ξ_f, η_f), or the death of a randomly chosen generating point is proposed. With some more details given in the Appendix, we mention here a few implementation issues specific to the current application.

The ground elevation surface ζ should retain its observed data values z_m at the sample points s_m, $m = 1, \ldots, M$. This is implemented by initialising $\boldsymbol{\xi}_\zeta$ to $\mathbf{s}$ and $\boldsymbol{\eta}_\zeta$ to $\mathbf{z}$, by never accepting a proposal where one tile contains two or more sample points s_m, and by always proposing level z_m to the tile that contains s_m. Some care is needed here to ensure reversibility (see Appendix for details). Since there are some adjacent sample points with identical elevation readings, one could alternatively reject only those proposals where one tile contains sample points with different observations.

The response function is, in principle, defined over the entire real line, which is somewhat problematic in our approach. For simplicity, we constrain the generating points $\xi_{\gamma,k}$ to lie inside the interval $\Delta_\gamma = [-1, 15]$ containing all observed elevations. Restrictions of the subintervals $E_k(\boldsymbol{\xi}_\gamma)$ to Δ_γ are applied whenever their lengths are needed in the dimension-changing moves (see Appendix). Note, however, that we do not restrict the sampled ground elevation values $\eta_{\zeta,k}$ to lie inside Δ_γ.

In order to facilitate the comparison to Rathbun's estimates, we used the scaling option that assumes that $\int_E \lambda_0(s)\,\nu(ds)/\nu(E) = 1$. This rescaling is performed periodically during the sampling (rather than afterwards), which helps avoid such drifts among unidentifiable parameters that could lead to numerical problems (cf., Besag et al., 1995, Section 4.1).

The results presented in Section 4 are based on samples of 10,000 realisations, collected by saving the current state after every 2000th basic update step after a burn-in period of 1,000,000 steps. This makes a total of 21,000,000 steps, which took about 10 hours on our workstation based on a 500-MHz

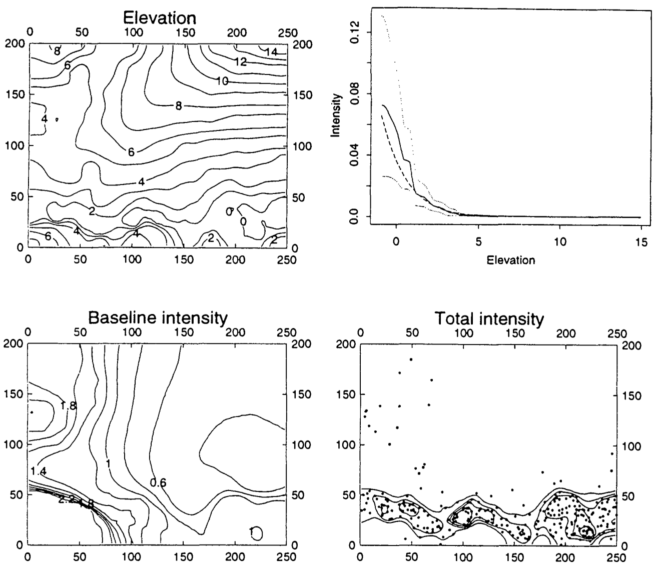

Figure 2. Pointwise posterior mean estimates for ironwood. For the response function γ (top right), Rathbun's log-quadratic estimate (dashed line) and our pointwise 80% credible intervals (dotted lines) are also shown. The observed point pattern is overlaid on the plot of total intensity.

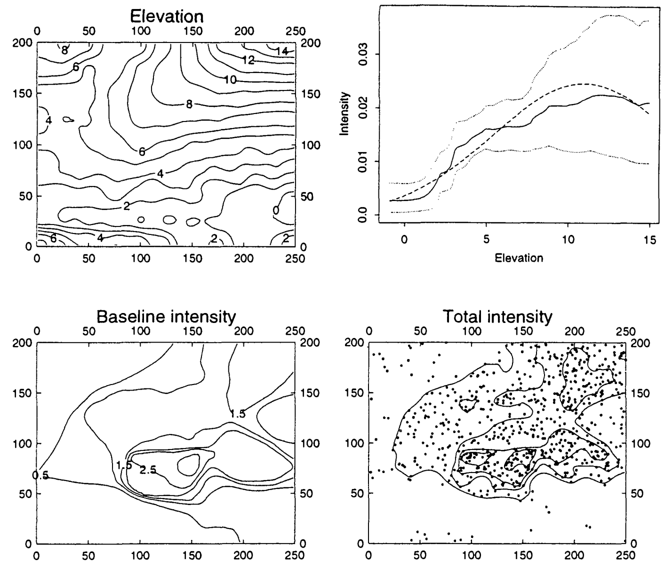

Figure 3. Pointwise posterior mean estimates for blue beech (as Figure 2).

DEC Alpha 21164 processor. The acceptance rates varied between 20 and 70%. Convergence was assessed by monitoring the likelihood $p(\mathbf{x} \mid \lambda)$, the sufficient statistics in the prior of each parameter function: K, $\Sigma \log w_{k+}$ and $\Sigma w_{kj}(\eta_k - \eta_j)^2$, and the values of the response curve γ at eight control points equally spaced on $[-1, 15]$. The Monte Carlo standard errors for the posterior means of these statistics were assessed via the initial monotone sequence estimator of Geyer (1992); they were all less than 13% of the corresponding posterior standard deviations.

The C-code of our sampler and some S-Plus functions for graphical display of the results are available over the Internet at `http://www.stat.jyu.fi/~jmhe/pub/Steps.html`.

4. Results

For an illustration of our method, we have selected two representative species, ironwood and blue beech *Ostrya virginiana*. Ironwood serves as an example of an 'easy' species, for which the parametric model of Rathbun (1996) gives good fit. Blue beech, on the other hand, is the species for which Rathbun reports worst fit. Figures 2 (ironwood) and 3 (blue beech) display curves and surfaces formed by our pointwise posterior mean estimates for elevation ζ, response curve γ, baseline intensity λ_0, and total intensity λ. For the response curve γ, the pointwise symmetric 80% credible intervals and the corresponding parametric estimate of Rathbun are also shown.

Consider first the topographic contour maps in the top left displays of Figures 2 and 3. Since we have treated ironwood and blue beech data separately, by repeating the same estimation procedure for both species, we have ended up producing two topographic contour maps for the Titi Hammock study region. This could be changed easily into a combined analysis by replacing the two point pattern likelihood expressions arising from single species by their product and using the same elevation values in both. However, the way the estimation has been done here, the two contour maps resemble each other closely and are also very similar to the map produced in Rathbun (1996, Figure 3).

For a comparison, Figure 4 shows yet another estimate of the elevation surface, this time without using the tree data at all, i.e., we have only used likelihood (2.3) with the same prior for ζ as in the other experiments. The estimate with blue beech data is almost identical to this 'prior' interpolation. The northern parts of 2- and 3-m contours at about 'latitude' $Y = 50$ are somewhat more wiggly, following apparent changes in the blue beech intensity. The effect of ironwood data on elevation estimates is more visible, especially in the southern valley, where ironwood is most abundant. Again, the

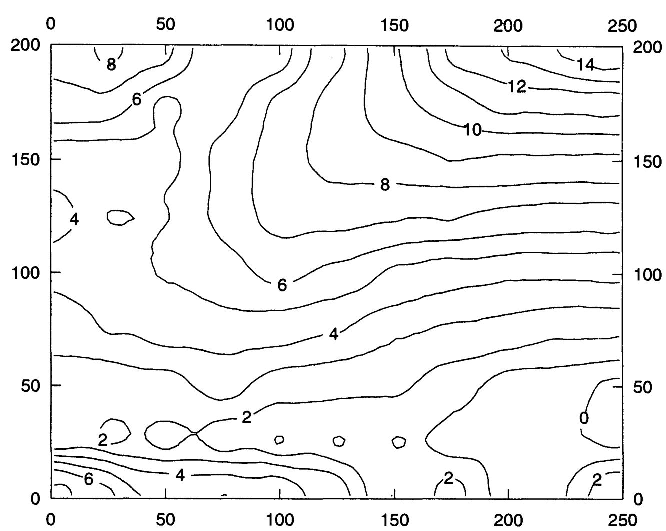

Figure 4. Posterior means of elevations without using any tree data.

topographic contours adjust to sudden changes in the tree intensity.

Consider then the estimated influence of elevation on ironwood and blue beech stands. Again, as shown in the top right displays of Figures 2 and 3, our pointwise posterior means of the function γ resemble rather well the corresponding parametric point estimates obtained by Rathbun. Apparently the most interesting observation here concerns blue beech: As already noted by Rathbun, his estimate for γ appears to be too high for elevations below 2 m. This does not happen in our nonparametric method, where the functional form of γ has not been specified in advance and where it is, in essence, data driven. We actually suspect that the bell-shaped exponential of a quadratic curve considered by Rathbun, resembling the normal density, does not give a particularly good description to the abundance of blue beech at higher elevations either.

The role of the baseline intensity λ_0 in the description of the Titi Hammock data is to explain the apparent extra-Poisson variability that would be present if the intensity were considered as a function of only the corresponding elevation. This element was missing from Rathbun's model completely. In the bottom left display of Figure 2, e.g., in the western part of the study area, the posterior mean of λ_0 assumes values as large as two or higher, and these are then balanced by much smaller values toward the east. Corresponding clusters of ironwood can be seen in the western part of the map, where there seems to be more trees growing than could be expected purely on grounds of the corresponding elevation readings. Similar comments apply to blue beech, where the baseline intensity peaks close to the centre of the study region (bottom left display of Figure 3).

For each unknown function f, there are two hyperparameters, $\lambda_{\boldsymbol{\xi},f}$ and β_f, whose values must be specified. The generating point intensity $\lambda_{\boldsymbol{\xi},f}$ has a direct interpretation as the prior mean number of generating points per unit length or area. In other words, it determines the typical tile size in the individual realisations and can be used to control how fine details of the function are shown. Its choice also has a considerable effect on the required computational effort, which increases with the number of tiles. In the experiments reported above, the values were chosen as follows:

- $\lambda_{\lambda_0} = 0.0005$, corresponding to prior mean number of 25 generating points for the baseline intensity surface in the 250-m × 200-m region;
- $\lambda_\zeta = 0.004$, corresponding to prior mean number of 200 generating points for the elevation surface;
- $\lambda_\gamma = 0.625$, corresponding to prior mean number of 10 generating points for the response curve.

A rough idea on the effect of β_f can be obtained by looking at the full conditionals (2.5) of the function value prior. Consider, e.g., the conditional prior distribution of $\eta_{\gamma,k}$, the value of $\log \gamma$ in $E_k(\boldsymbol{\xi}_\gamma)$, given all other values and assuming, for simplicity, that $w_{k+} = 1$, which is a rather typical value with the current choice of λ_γ. Given τ_γ, the standard deviation of this distribution is $\sigma = \tau_\gamma^{-1/2}$. The prior densities of σ corresponding to three distinct values of β_γ are shown in the left-hand display of Figure 5. Note that difference $\log 2 \approx 0.7$ in η_γ values corresponds to halving or doubling the intensity. The value of β_γ applied in the reported experiments was 0.05.

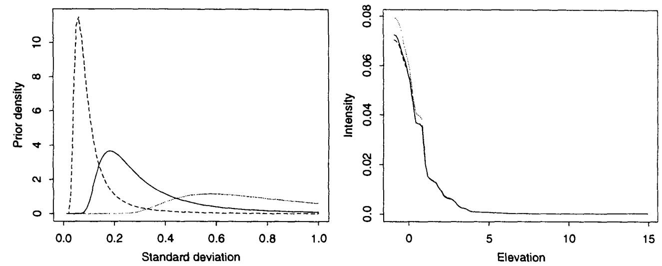

Figure 5. **Left.** Prior densities of $\tau_\gamma^{-1/2}$ (standard deviation of the conditional distribution (2.5) of $\eta_{\gamma,k}$, when $w_{k+} = 1$) with $\beta_\gamma = 0.005$ (dashed line), $\beta_\gamma = 0.05$ (solid line), and $\beta_\gamma = 0.5$ (dotted line). **Right.** Corresponding posterior mean estimates of response function γ for ironwood.

The right-hand display of Figure 5 indicates that the results are not very sensitive to this choice. For $f \in \{\lambda_0, \zeta\}$, β_f was set to 0.01.

5. Discussion

In this paper, we have proposed a nonparametric intensity model for describing spatial point pattern data and for relating such a description to the values of a concomitant variable from which only point measurements are assumed to be available. The suggested model structure involves two nonparametrically defined spatial (bivariate) functions, one representing the true values of the considered concomitant variable and the other serving as a baseline intensity, describing observed extra-Poisson residual variation. A third (univariate) nonparametric function is used as a link between the concomitant variable and the observed point pattern response.

The suggested statistical inference from this model follows a fully Bayesian estimation scheme, where all three nonparametric functions are estimated jointly. In practice, the numerical work involves the application of computationally intensive algorithmic Markov chain Monte Carlo methods. Their implementation is not a routine task, and rigid convergence assessment is inevitably difficult. Serious nonconvergence, however, is usually detected by simple diagnostics.

We believe that these general modeling and inferential ideas, involving a combination of several nonparametrically defined functions to form a likelihood expression and their joint Bayesian estimation from observed data, have a much greater potential in applications than we have been able to illustrate with our concrete example. Response surfaces other than the Poisson intensity could be modeled simply by choosing a different likelihood expression; restoration of a grey level image can be viewed as a particular case of such a task. Image classification is an example of the case where the function of interest is truly piecewise constant and its values restricted to a (small) finite set of labels. In such a case, we could replace the Gaussian function value prior by, e.g., the Potts model.

ACKNOWLEDGEMENTS

We are extremely grateful to Steve Rathbun for kindly providing us with the Titi Hammock data. This work was partly supported by a research grant from the Academy of Finland. Computing facilities of the Department of Statistics, University of Jyväskylä, were used. Rolf Turner's ratfor routines from StatLib-archive (`http://lib.stat.cmu.edu/general/delaunay`) were modified to perform Voronoi tessellations. The insightful comments by the editor and two referees have helped us improve the manuscript.

Résumé

Nous présentons une formulation bayésienne non-paramétrique du problème de modélisation de processus spatiaux ponctuels influencés par des variables concomitantes. L'information sur ces variables est supposée incomplète, consistant en un nombre relativement faible de points de mesure. La variation résiduelle, causée par d'autres facteurs influants non mesurés est modélisée en termes d'une intensité de base variable dans l'espace. Un schéma de simulation de Monte-Carlo par chaînes de Markov est proposé pour l'estimation non-paramétrique simultanée de chacune des fonctions inconnues du modèle. La méthode proposée est illustrée par une réanalyse des données de Rathbun (1996, *Biometrics* **52,** 226–242); les modèles estimés sont comparés à ceux obtenus par Rathbun.

References

Arjas, E. and Andreev, A. (1996). A note on histogram approximation in Bayesian density estimation. In *Bayesian Statistics,* Volume 5, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds), 487–490. Oxford: Oxford University Press.

Arjas, E. and Heikkinen, J. (1997). An algorithm for nonparametric Bayesian estimation of a Poisson intensity. *Computational Statistics* **12,** 385–402.

Besag, J., Green, P. J., Higdon, D., and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science* **10,** 3–66.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo (with discussion). *Statistical Science* **7,** 473–511.

Green, P. J. (1995). Reversible jump MCMC and Bayesian model determination. *Biometrika* **82,** 711–732.

Heikkinen, J. (1998). Curve and surface estimation using dynamic step functions. In *Practical Nonparametric and Semiparametric Bayesian Statistics,* D. Dey, P. Müller, and D. Sinha (eds), 255–272. New York: Springer-Verlag.

Heikkinen, J. and Arjas, E. (1998). Non-parametric Bayesian estimation of a spatial Poisson intensity. *Scandinavian Journal of Statistics* **25,** 435–450.

Rathbun, S. L. (1996). Estimation of Poisson intensity using partially observed concomitant variables. *Biometrics* **52,** 226–242.

Received November 1997. *Revised July* 1998.
Accepted September 1998.

Appendix

Formulas for the Posterior Sampler

This appendix gives the formulas applied in the reversible jump MCMC sampling from the posterior distribution. For more details on their derivation, the reader is referred to Heikkinen (1998).

Let $\boldsymbol{\theta} = (\tau_{\lambda_0}, \boldsymbol{\xi}_{\lambda_0}, \boldsymbol{\eta}_{\lambda_0}, \tau_\zeta, \boldsymbol{\xi}_\zeta, \boldsymbol{\eta}_\zeta, \tau_\gamma, \boldsymbol{\xi}_\gamma, \boldsymbol{\eta}_\gamma)$ denote the complete set of our model parameters. The joint posterior $p(\boldsymbol{\theta} \mid \mathbf{x}, \mathbf{z})$ of $\boldsymbol{\theta}$ is proportional to the expression

$$\left\{ \prod_{f=\lambda_0,\zeta,\gamma} p(\tau_f) p(\boldsymbol{\xi}_f) p(\boldsymbol{\eta}_f \mid \boldsymbol{\xi}_f, \tau_f) \right\} p(\mathbf{x} \mid \boldsymbol{\theta}) p(\mathbf{z} \mid \boldsymbol{\theta}),$$

where the Poisson likelihood function can be written as

$$\begin{aligned} p(\mathbf{x} \mid \boldsymbol{\theta}) \propto \exp\Bigg[& \sum_{k=1}^{K_{\lambda_0}} \eta_{\lambda_0,k} N_{\mathbf{x}}\{E_k(\boldsymbol{\xi}_{\lambda_0})\} \\ & + \sum_{k=1}^{K_\zeta} \eta_{\gamma,i(\eta_{\zeta,k})} N_{\mathbf{x}}\{E_k(\boldsymbol{\xi}_\zeta)\} \\ & - \sum_{k=1}^{K_{\lambda_0}} \sum_{j=1}^{K_\zeta} \exp(\eta_{\lambda_0,k} + \eta_{\gamma,i(\eta_{\zeta,j})}) \\ & \times \nu\{E_k(\boldsymbol{\xi}_{\lambda_0}) \cap E_j(\boldsymbol{\xi}_\zeta)\} \Bigg]. \end{aligned}$$

Here $N_{\mathbf{x}}(A)$ is the number of individual points in pattern $\mathbf{x}$ inside domain A, $\nu(A)$ is the size (Lebesgue measure) of A (i.e., the length in one dimension and the area in two), and $i(z)$ is the index of the interval in $\mathcal{E}(\boldsymbol{\xi}_\gamma)$ that contains z. For future reference, note the two following alternative ways of rewriting the double sum:

$$\begin{aligned} & \sum_{k=1}^{K_{\lambda_0}} \sum_{j=1}^{K_\zeta} \exp(\eta_{\lambda_0,k} + \eta_{\gamma,i(\eta_{\zeta,j})}) \nu\{E_k(\boldsymbol{\xi}_{\lambda_0}) \cap E_j(\boldsymbol{\xi}_\zeta)\} \\ & = \sum_{k=1}^{K_{\lambda_0}} \left[\exp \eta_{\lambda_0,k} \sum_j \exp(\eta_{\gamma,i(\eta_{\zeta,j})}) \nu\{E_j(\boldsymbol{\xi}_\zeta) \cap E_k(\boldsymbol{\xi}_{\lambda_0})\} \right] \\ & = \sum_{k=1}^{K_\zeta} \left[\exp \eta_{\gamma,i(\eta_{\zeta,k})} \sum_j \exp(\eta_{\lambda_0,j}) \nu\{E_j(\boldsymbol{\xi}_{\lambda_0}) \cap E_k(\boldsymbol{\xi}_\zeta)\} \right]. \end{aligned}$$

The inner sums typically contain only a few terms since most intersections are empty.

In each iteration of the posterior sampler, one of the following moves is proposed for a randomly selected parameter function $f \in \{\lambda_0, \zeta, \gamma\}$:

(1) Change the precision parameter value τ_f.
(2) Change one randomly chosen function value $\eta_{f,k}$, $k \in \{1, \ldots, K_f\}$.
(3) Add a new (marked) generating point (ξ'_f, η'_f).
(4) Remove one randomly chosen generating point $\xi_{f,k}$, $k \in \{1, \ldots, K_f\}$.

To simplify notation, the subscript f will be omitted below whenever there is no danger of confusion. The proposed moves from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}'$ are accepted with probability

$$\min\left\{1, \frac{q(\boldsymbol{\theta}' \to \boldsymbol{\theta})}{q(\boldsymbol{\theta} \to \boldsymbol{\theta}')} \frac{p(\boldsymbol{\theta}' \mid \mathbf{x}, \mathbf{z})}{p(\boldsymbol{\theta} \mid \mathbf{x}, \mathbf{z})}\right\}, \qquad \text{(A.1)}$$

where q is the density of the proposal probability kernel. Our dimension-changing moves are such that the Jacobian (Green, 1995) is always equal to one, and therefore it is omitted in expression (A.1).

For a type 1 move, we create the proposal τ' by drawing $\log\tau'$ from the uniform distribution on $[\log\tau - C_\tau, \log\tau + C_\tau]$, where C_τ (as well as other C's appearing later) is a sampler parameter that can be tuned to improve mixing. The corresponding proposal ratio is $q(\tau' \to \tau)/q(\tau \to \tau') = \tau'/\tau$ due to the log-transform. The posterior ratio is

$$\frac{p(\tau')p(\boldsymbol{\eta} \mid \boldsymbol{\xi}, \tau')}{p(\tau)p(\boldsymbol{\eta} \mid \boldsymbol{\xi}, \tau)} = (\tau'/\tau)^{K/2} \exp\left[-\left\{\beta + \frac{1}{2}\sum_{k<j} w_{kj}(\eta_k - \eta_j)^2\right\}(\tau' - \tau)\right].$$

In a type 2 move, an index k is sampled from the uniform distribution on integers $1, \ldots, K_f$. If $f = \zeta$ and $E_k(\boldsymbol{\xi}_\zeta) \cap \mathbf{s} \neq \emptyset$, we refuse this proposal without further consideration since any change to $\eta_{\zeta,k}$ would then lead to $p(\mathbf{z} \mid \boldsymbol{\theta}) = 0$. Otherwise, we draw a proposal η'_k from the uniform distribution on the interval $[\eta_k - C_\eta, \eta_k + C_\eta]$. The proposal ratio cancels due to the symmetry, and the posterior ratio is

$$\frac{p(\eta'_k \mid \boldsymbol{\eta}_{-k}, \boldsymbol{\xi}, \tau)p(\mathbf{x} \mid \boldsymbol{\theta}')}{p(\eta_k \mid \boldsymbol{\eta}_{-k}, \boldsymbol{\xi}, \tau)p(\mathbf{x} \mid \boldsymbol{\theta}')} = \exp\left[-\tau(\eta'_k - \eta_k)\left\{w_{k+}(\eta'_k + \eta_k)/2 - \sum_j w_{kj}\eta_j\right\} + D_{\text{like},1}\right],$$

where $D_{\text{like},1} = D_{\text{like},f,1} = \log p(\mathbf{x} \mid \boldsymbol{\theta}') - \log p(\mathbf{x} \mid \boldsymbol{\theta})$ is the log-likelihood difference

$$\begin{aligned} D_{\text{like},\lambda_0,1} = {} & (\eta'_{\lambda_0,k} - \eta_{\lambda_0,k})N_{\mathbf{x}}\{E_k(\boldsymbol{\xi}_{\lambda_0})\} \\ & - (\exp\eta'_{\lambda_0,k} - \exp\eta_{\lambda_0,k}) \\ & \times \sum_j \exp(\eta_{\gamma,i(\eta_{\zeta,j})})\nu\{E_j(\boldsymbol{\xi}_\zeta) \cap E_k(\boldsymbol{\xi}_{\lambda_0})\}, \\ D_{\text{like},\zeta,1} = {} & \{\eta_{\gamma,i(\eta'_{\zeta,k})} - \eta_{\gamma,i(\eta_{\zeta,k})}\}N_{\mathbf{x}}\{E_k(\boldsymbol{\xi}_\zeta)\} \\ & - (\exp\eta_{\gamma,i(\eta'_{\zeta,k})} - \exp\eta_{\gamma,i(\eta_{\zeta,k})}) \\ & \times \sum_j \exp(\eta_{\lambda_0,j})\nu\{E_j(\boldsymbol{\xi}_{\lambda_0}) \cap E_k(\boldsymbol{\xi}_\zeta)\}, \\ D_{\text{like},\gamma,1} = {} & \{\eta'_{\gamma,k} - \eta_{\gamma,k}\} \sum_{j:i(\eta_{\zeta,j})=k} N_{\mathbf{x}}\{E_j(\boldsymbol{\xi}_\zeta)\} \\ & - (\exp\eta'_{\gamma,k} - \exp\eta_{\gamma,k}) \\ & \times \sum_{j:i(\eta_{\zeta,j})=k} \sum_i \exp(\eta_{\lambda_0,i})\nu\{E_i(\boldsymbol{\xi}_{\lambda_0}) \cap E_j(\boldsymbol{\xi}_\zeta)\}. \end{aligned}$$

Move types 3 and 4 are designed to form pairs of reversible jumps between different dimensions. To fix the notation, let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_K)$ and $\boldsymbol{\xi}' = (\xi'_1, \ldots, \xi'_{K'})$, where $K' = K + 1$ and $\xi'_k = \xi_k$ for $k = 1, \ldots, K$, i.e., $\boldsymbol{\xi}'$ is obtained by adding one point to $\boldsymbol{\xi}$ and indexing it as the last one or, reversely, $\boldsymbol{\xi}$ is obtained by deleting $\xi'_{K'}$ from $\boldsymbol{\xi}'$. Define $\boldsymbol{\eta}$ and $\boldsymbol{\eta}'$ analogously and consider a pair $\boldsymbol{\theta}$, $\boldsymbol{\theta}'$ of states, identical except for these differences in the currently updated function f.

In a birth proposal from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}'$, the proposed location of a new generating point $\xi'_{K'}$ is drawn from the uniform distribution on Δ_f, where $\Delta_f = E$ for $f \in \{\lambda_0, \zeta\}$. Let $\mathcal{K}'$ denote the set $\{k : \xi'_k \underset{\boldsymbol{\xi}'}{\sim} \xi'_{K'}\}$ of indices in the neighbourhood of the new point. If $f \in \{\lambda_0, \gamma\}$ or $f = \zeta$ and the new tile $E'_{\zeta,K'_\zeta}(\boldsymbol{\xi}'_\zeta)$ does not contain any of the sample points $\mathbf{s}$, then we propose to mark $\eta'_{K'} = \tilde{\eta} + \varepsilon$, where

$$\tilde{\eta} = \sum_{k\in\mathcal{K}'} \frac{\nu\{E_k(\boldsymbol{\xi})\} - \nu\{E_k(\boldsymbol{\xi}')\}}{\nu\{E_{K'}(\boldsymbol{\xi}')\}}\eta_k$$

is a weighted average of the current function values in the neighbouring tiles and perturbation $\varepsilon \in \mathbb{R}$ is drawn from the density $g(\varepsilon) = C_\varepsilon \exp(C_\varepsilon\varepsilon)/\{1 + \exp(C_\varepsilon\varepsilon)\}^2$. (For $f = \gamma$, the intervals $E_k(\boldsymbol{\xi}_\gamma)$ are restricted to Δ_γ when measuring their lengths ν.) The proposal ratio for this pair of states is

$$\frac{q(\boldsymbol{\theta}' \to \boldsymbol{\theta})}{q(\boldsymbol{\theta} \to \boldsymbol{\theta}')} = \frac{d_{K'}/K'}{b_K g(\eta'_{K'} - \tilde{\eta})/\nu(\Delta)},$$

where b_k and d_k denote the probabilities with which birth and death moves, respectively, are proposed when the current number of generating points is k. We have chosen them to satisfy equations $d_{k+1}/b_k = (k+1)/\lambda_{\boldsymbol{\xi}}\nu(\Delta)$, whereby the proposal ratio simplifies to $\{\lambda_{\boldsymbol{\xi}} g(\eta'_{K'} - \tilde{\eta})\}^{-1}$. Letting w' denote a weight derived from pattern $\boldsymbol{\xi}'$, the posterior ratio can be expressed as

$$\frac{p(\boldsymbol{\xi}')p(\boldsymbol{\eta}' \mid \boldsymbol{\xi}', \tau)p(\mathbf{x} \mid \boldsymbol{\theta}')}{p(\boldsymbol{\xi})p(\boldsymbol{\eta} \mid \boldsymbol{\xi}, \tau)p(\mathbf{x} \mid \boldsymbol{\theta})} = \lambda_{\boldsymbol{\xi}}\left(\frac{\tau_f w'_{K'+}}{2\pi}\prod_{k\in\mathcal{K}'}\frac{w'_{k+}}{w_{k+}}\right)^{1/2} \times \exp\left(-\frac{\tau_f}{2}D_{\text{prior}} + D_{\text{like},2}\right),$$

where

$$\begin{aligned} D_{\text{prior}} &= \sum_{k\in\mathcal{K}'}\left[w'_{K'k}(\eta'_{K'} - \eta_k)^2 - \frac{1}{2}\sum_{j\in\mathcal{K}'}\{(w'_{kj} - w_{kj})(\eta_k - \eta_j)^2\}\right], \\ D_{\text{like},\lambda_0,2} &= \sum_{k\in\mathcal{K}'}\Big[(\eta'_{\lambda_0,K'} - \eta_{\lambda_0,k})N_{\mathbf{x}}\{E_{K'}(\boldsymbol{\xi}'_{\lambda_0}) \cap E_k(\boldsymbol{\xi}_{\lambda_0})\} \\ &\quad - (\exp\eta'_{\lambda_0,K'} - \exp\eta_{\lambda_0,k}) \times \sum_j \exp(\eta_{\gamma,i(\eta_{\zeta,j})}) \\ &\quad \times \nu\{E_j(\boldsymbol{\xi}_\zeta) \cap E_{K'}(\boldsymbol{\xi}'_{\lambda_0}) \cap E_k(\boldsymbol{\xi}_{\lambda_0})\}\Big], \\ D_{\text{like},\zeta,2} &= \sum_{k\in\mathcal{K}'}\Big[(\eta_{\gamma,i(\eta'_{\zeta,K'})} - \eta_{\gamma,i(\eta_{\zeta,k})})N_{\mathbf{x}}\{E_{K'}(\boldsymbol{\xi}'_\zeta) \cap E_k(\boldsymbol{\xi}_\zeta)\} \\ &\quad - (\exp\eta_{\gamma,i(\eta'_{\zeta,K'})} - \exp\eta_{\gamma,i(\eta_{\zeta,k})}) \times \sum_j \exp(\eta_{\lambda_0,j}) \\ &\quad \times \nu\{E_j(\boldsymbol{\xi}_{\lambda_0}) \cap E_{K'}(\boldsymbol{\xi}'_\zeta) \cap E_k(\boldsymbol{\xi}_\zeta)\}\Big], \end{aligned}$$

$$D_{\text{like},\gamma,2} = \sum_{k:i(\eta_{\zeta,k})=K'} \Big[(\eta'_{\gamma,K'} - \eta_{\gamma,i(\eta_{\zeta,k})}) N_{\mathbf{x}}\{E_k(\boldsymbol{\xi}_\zeta)\} - (\exp \eta'_{\gamma,K'} - \exp \eta_{\gamma,i(\eta_{\zeta,k})}) \times \sum_j \exp(\eta_{\lambda_0,j}) \nu\{E_j(\boldsymbol{\xi}_{\lambda_0}) \cap E_k(\boldsymbol{\xi}_\zeta)\} \Big].$$

If $f = \zeta$ and $\mathbf{s} \cap E_{\zeta,K'_\zeta}(\boldsymbol{\xi}'_\zeta) = \{s_m\}$, then we propose $\eta'_{\zeta,K'_\zeta} = z_m$ so that $\zeta(s_m)$ remains equal to z_m and the likelihood $p(\mathbf{z} \mid \zeta)$ remains positive. In order to obtain pairs of reversible jumps, we also propose a perturbation $\eta'_{\zeta,k} = \eta_{\zeta,k} + \varepsilon (= z_m + \varepsilon)$ to the tile $E_k(\boldsymbol{\xi}_\zeta)$ that contains s_m in the current tessellation; as earlier, ε is drawn from density $g(\varepsilon)$. This birth move is reversed by removing $\xi'_{K'}$ and setting $\eta_{\zeta,k}$ to z_m. The proposal ratio is as above but with g evaluated at $\eta'_{\zeta,k} - \eta_{\zeta,k}$. The expressions for D_{prior} and $D_{\text{like},\zeta,2}$ become a bit more complicated since $\eta'_{\zeta,k} \neq \eta_{\zeta,k}$.

Finally, if $f = \zeta$ and the proposed new tile $E_{\zeta,K'_\zeta}(\boldsymbol{\xi}'_\zeta)$ would contain more than one of the sample points $\mathbf{s}$, then we automatically refuse that proposal. Also, death proposals with two sample points in one tile $E_{\zeta,k}(\boldsymbol{\xi}_\zeta)$ are automatically refused.