Bayesian analysis of log Gaussian Cox processes for disease mapping

Viktor Beneš[*†] Karel Bodlák[†] Jesper Møller[‡]
Rasmus Waagepetersen[‡]

Abstract

We consider a data set of locations where people in Central Bohemia have been infected by tick-borne encephalitis, and where population census data and covariates concerning vegetation and altitude are available. The aims are to estimate the risk map of the disease and to study the dependence of the risk on the covariates. Instead of using the common area level approaches we consider a Bayesian analysis for a log Gaussian Cox point process with covariates. Posterior characteristics for a discretized version of the log Gaussian Cox process are computed using Markov chain Monte Carlo methods. A particular problem is to determine a model for the population intensity, and the dependence of the results on the model for the population intensity is discussed in detail. Model validation is based on the posterior predictive distribution of various summary statistics.

Keywords: population intensity, Langevin-Hastings algorithm, Markov chain Monte Carlo, spatial point process.

1 Introduction

The aims of statistical disease mapping are to characterize the spatial variation of cases of a disease and to investigate connections with possible covariates. In particular it is of interest to identify areas with an elevated disease risk. The data may be a point pattern showing e.g. home residences of diseased people or locations where people have acquired an infection. Often, the data are aggregated

[*]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 18208 Praha 8, Czech Republic

[†]Department of Probability and Statistics, Charles University, Sokolovska 83, 18675 Praha 8, Czech Republic

[‡]Department of Mathematical Sciences, Aalborg University, Fredrik Bajersvej 7G, DK-9220 Aalborg, Denmark

so that only counts of diseased people within subregions of the study region are available. Indeed, most statistical analyses reported in the literature are based on a so-called area level approach, where a model for aggregated data is used after an initial aggregation. However, if a spatial point pattern is available, it is more natural to use a spatial point process model. Recent surveys of both the area level approach and point process modelling in disease mapping are given by Diggle (2000), Richardson (2001), and the accompanying discussions by Knorr-Held (2001) and Møller (2001). For disease mapping in general, see Best *et al.* (1999), Lawson *et al.* (2001), and the references therein.

In this paper we consider a point process approach to the analysis of a data set of locations where people in Central Bohemia have been infected by tick-borne encephalitis (TBE). Specifically we consider a log Gaussian Cox point process (LGCP), where covariates concerning occurrence of different forest types, altitude, and the population density are used in the modelling of the spatially varying intensity of TBE infections. LGCPs were independently introduced in astronomy by Coles and Jones (1991) and in statistics by Møller *et al.* (1998).

A particular problem is the determination of the population intensity of humans being at risk. Raw geographical population data connects population numbers to home locations, but typically people get infected during visits to more or less distant surroundings of their homes. This is an additional complication compared with spatial analysis of chronic diseases like cancer, where the objective may be to study association between disease incidence and risk factors at the home locations. We consider various approaches to smoothing of population data, where the smoothing to some extent is related to the movement of people around their homes.

The data and previous analyses in Zeman (1997) and other papers are described in more detail in Section 2. Section 3 considers estimation of the population intensity, modelling of the risk function in terms of a LGCP depending on covariates, and our Bayesian approach to inference using Markov chain Monte Carlo methods. The results for different models of the population intensity are discussed in Section 4.

2 Data and background

2.1 Description of data and scientific problem

TBE is an infectious debilitating illness which is transmitted by parasitic ticks and which occasionally afflicts humans. Epidemiologists and medical practitioners making decision on prophylactic measures deal with the problem of estimating the risk that a human gets infected by TBE at a specific location, cf. Zeman (1997). Since field studies of potential animal hosts are expensive, usually the data for statistical analysis consist of case locations and a population map. Moreover, explanatory variables of geographical nature which may influence the risk of infection are often available from geographical information systems.

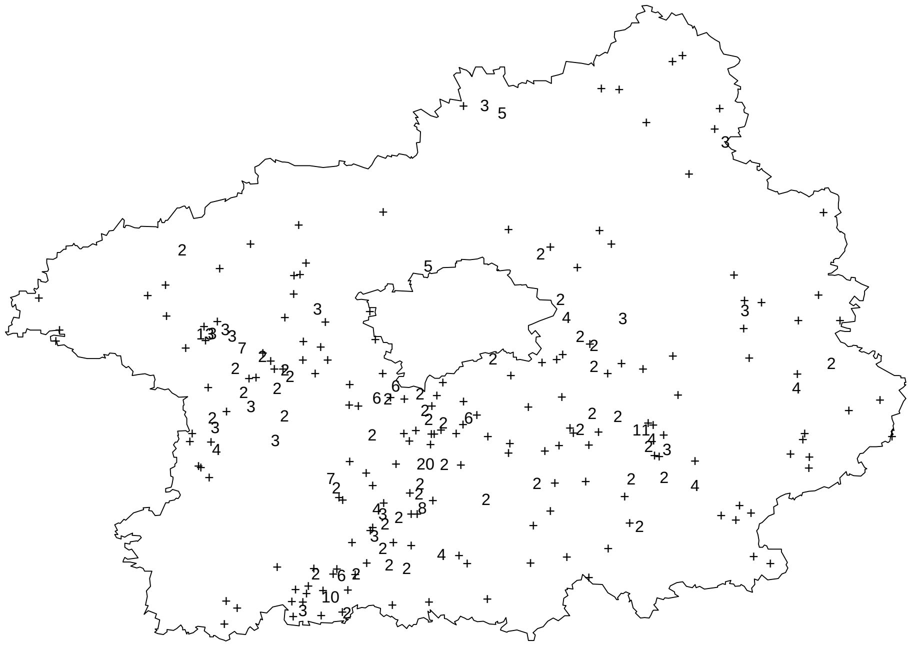

Figure 1: Locations of infection of 446 cases of tick-borne encephalitis in Central Bohemia. For each distinct location the number of cases associated with the location is shown (a plus corresponds to one case).

Figure 1 shows the point pattern of locations of 446 cases of TBE in Central Bohemia reported during 1971-93 by inhabitants living in Central Bohemia. The empty space in the middle of the figure corresponds to the capital Prague, and the total area of Central Bohemia is about 11860 km^2. This data set was first studied in Zeman (1997). Only 255 distinct points are visible due to ties in the data caused by positional error where several cases in an area have been associated with a common representative point. The distinct points in Figure 1 are marked with the number of cases associated with each point. Information concerning the distribution of the positional error is not available.

Let S denote the region of Central Bohemia. We have six covariates $d_1(s), \ldots,$ $d_6(s)$, $s \in S$, which are shown in Figures 2a–d. Here $d_1(\cdot), \ldots, d_5(\cdot)$ are indicator functions for the subareas of S covered by small forests (area between 0.1 and 0.5 km^2), medium forests (area between 0.5 and 1.5 km^2), coniferous forest, mixed

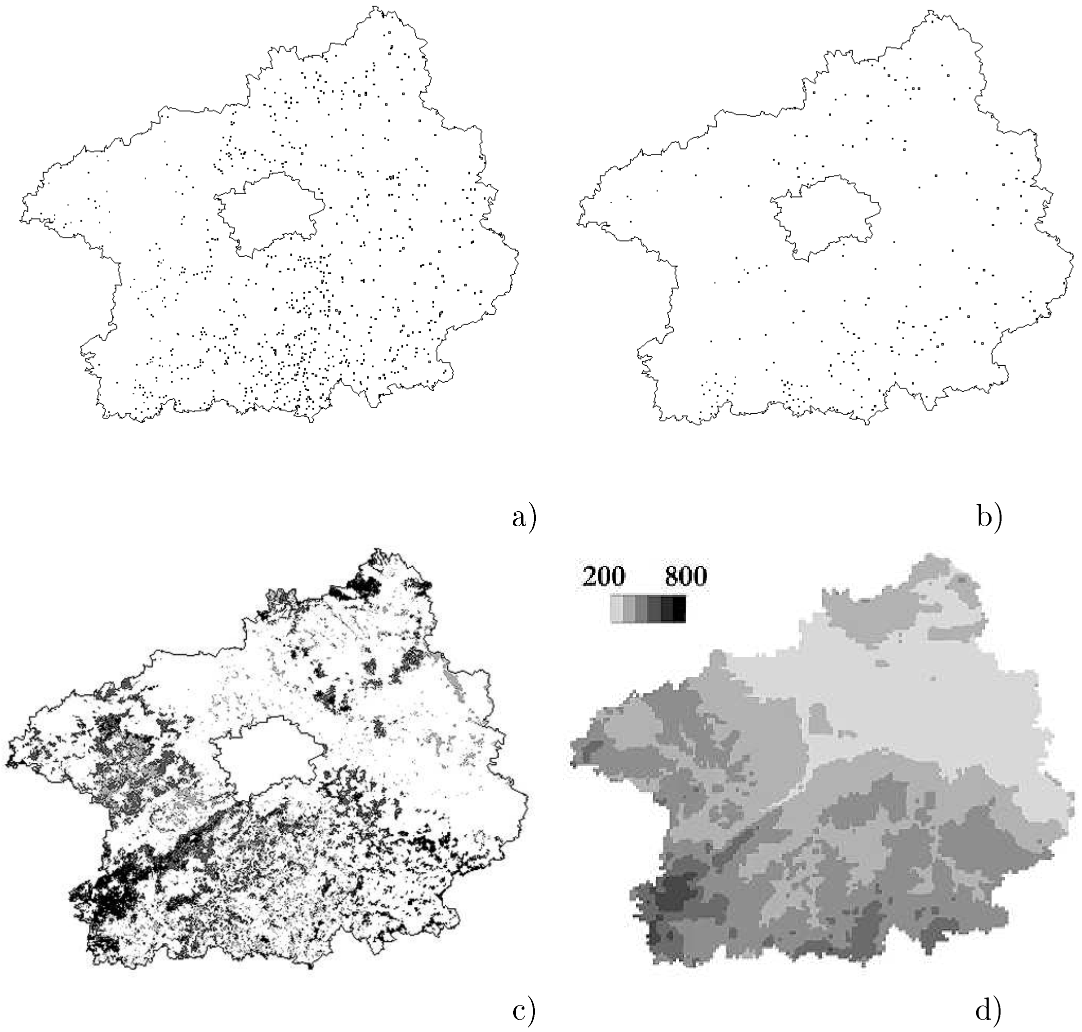

Figure 2: a) Small forests (0.1-0.5 km^2) and b) medium forests (0.5-1.5 km^2) (independent of the forest type). c) Three types of forest. Conifer: black; mixed: dark grey; foliate: light grey. d) Map of altitudes (in metres).

forest, and foliate forest, respectively, and $d_6(\cdot)$ is altitude in km. Each small or medium forest is represented by a disc of area equal to the area of the forest, and these covariates are possibly relevant since ticks can be transmitted by deers and other animals living in small or medium size forest areas. The covariates $d_i(\cdot)$, $i = 1, \ldots, 5$, are obtained from a LANDSAT-5 MSS satellite image of resolution 80×80 m^2. The covariate $d_6(\cdot)$ is obtained from the Institute of Military Topography, Dobruška.

Finally, population data from the National Census Bureau, Prague, are available. For the Central Bohemia they consist of the number of inhabitants in 3582 administrative units. In Figure 3 each unit is represented by a disc with center at a census point and radius given by $0.02\sqrt{\#\text{inhabitants in the unit}}$ km. Clusters of discs correspond to larger towns and cities. The total number of inhabitants is 1,112,717 and the largest city has about 74,000 inhabitants.

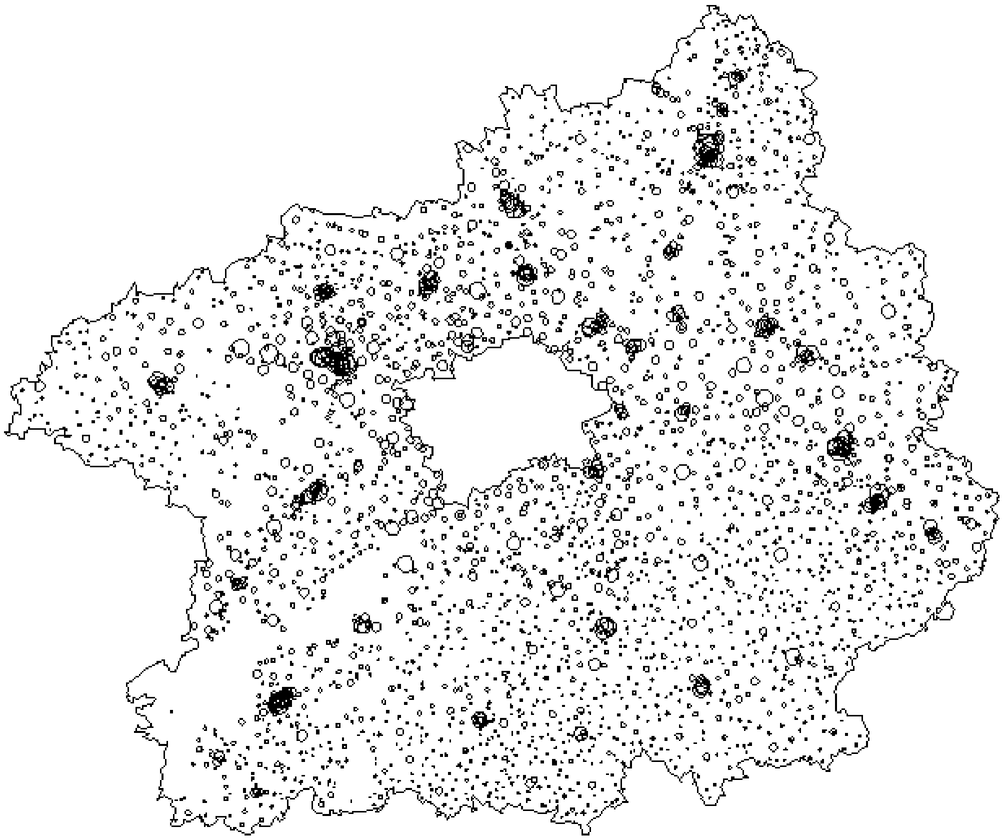

Figure 3: Population at 3582 administrative units in Central Bohemia represented by discs (see the text).

2.2 Previous data analysis

Zeman (1997) considers both the point pattern of TBE cases and another point pattern of cases for a related disease, Lyme borreliosis (LB). The LB data consist of paired data: 866 reported locations of infection during 1987-91 in Central Bohemia and the home location of each infected person. Zeman (1997) uses the distances between cases of infection and home location to obtain a kernel for smoothing the population map; see Section 3.2 also. Apart from this smoothed population map no other covariates are included in Zeman's analysis where the intensity functions of TBE and LB cases (each considered as a realization of a point process) are estimated by kernel methods. For each disease, Zeman

(1997) obtains a risk map by the ratio of the estimated intensity function and the smoothed population map (Bithell, 1990).

A similar ratio estimator of the risk map is suggested by Krejčíř (2000) where again both the TBE and the LB data are analyzed, using only the population data as an explanatory variable. He assumes that each point pattern of cases is a realization of an inhomogeneous Poisson point process with an intensity function constructed from beta splines, where the coefficients are estimated by a maximum likelihood method.

Incorporating the other explanatory variables in the model has so far only been studied in connection to two area level approaches for the TBE data. Mašata (1999) divides Central Bohemia into 41 irregular subregions, and he includes three covariates (the area in percentage of conifer, mixed, and foliate forests in each subregion) in a Bayesian Gaussian-Gaussian model (Stern and Cressie, 1999). Jiruše *et al.* (2000) use a subdivison of 141 squares of 10×10 km^2, and include the same types of covariates as Mašata (1999) together with the mean altitude and the proportion covered by small forests in each square. They use first a generalized linear model and the Akaike Information Criterion to find an optimal set of parameters, and second an empirical Bayesian approach to estimate the risk. Jiruše *et al.* (2000) compare the credibility intervals for risk estimators obtained by their method with those of Mašata (1999), and conclude that rather similar results are obtained in subregions with a large risk for infection, although it is only the model in Mašata (1999) which incorporates spatial dependence. The results of the above-mentioned papers are further discussed in Section 4.

3 Bayesian analysis using log Gaussian Cox processes

We use the following hierarchical model: At the first level, the point pattern x of locations of tick cases is assumed to be a realization of a Poisson process X with an intensity function which is a product of a population intensity and a risk function as described in Section 3.1. Estimation of the population intensity is discussed in Section 3.2. At the next level a log linear model for the risk function is proposed in Section 3.3, incorporating the covariate information and a Gaussian process which takes account of spatial variation not explained by the covariates and the uncertainty of the estimated population intensity. At the final stage hyper priors on the unknown parameters for the covariates and the Gaussian field are imposed, whereby a posterior is obtained in Section 3.4. For computational reasons certain approximations of the posterior are required as discussed in Section 3.5 and the Appendix. Finally, Section 3.6 deals with MCMC simulation of the posterior.

Strictly speaking, the multiple points in x cannot occur under the proposed model. However, our approximate approach actually only utilize counts of locations within certain small cells and this makes the results less sensitive to the presence of ties in x.

3.1 A simplified model

Our modelling of the TBE data is motivated by the following simplifying considerations, which are similar to one of the steps in the construction of a Neyman-Scott process (Neyman and Scott, 1958; Diggle, 1983).

In the observation period 1971-93 a number $m = 1,112,717$ of persons are living at home locations $h_1, \dots, h_m \in S$, and the ith person makes a number N_i of visits to the surroundings of h_i. The N_i's are assumed to be independent and Poisson distributed with mean $\lambda > 0$ independent of i. Given the N_i, the location of each visit of the ith person is distributed according to some density g_{h_i} on S, and the locations of visits of all persons are assumed to be independent. For a visit to a location $s \in S$, there is associated a probability $\pi(s)$ for getting an infection during the visit. The random set of locations where persons have been infected is then a Poisson process with intensity function of the form

$$\Lambda(s) = \rho(s)\pi(s), \ s \in S \tag{1}$$

where $\rho(s) = \lambda \sum_{i=1}^{m} g_{h_i}(s)$ is the *population intensity* of humans visiting s.

3.2 Estimation of population intensity

The population intensity $\rho(s)$ is a crucial component of the modelling. As it is unknown, we discuss below how it may be estimated; see also the discussion in Section 4.

For the LB data both locations of infection and home are available. Under the crude assumption that the densities g_{h_i} are of the form $g_{h_i}(s) = g(\|s - h_i\|)$ one may as in Zeman (1997) try to estimate g from the LB data. Recall that if f denotes the density of $\|Z\|$ for a two-dimensional random variable Z with isotropic density $g_h(z) = g(\|z - h\|)$, then g and f are related by

$$g(\|z\|) = f(\|z\|)/(2\pi\|z\|), \quad z \in \mathbb{R}^2. \tag{2}$$

Zeman (1997) fits a power regression to a histogram for the log distances between home and place of infection. He then obtains an expression $\tilde{f}(h) = ah^b$ for the density of the distances and uses this as a kernel for smoothing of the population data. Strictly speaking $\tilde{f}$ is not a proper density on $\mathbb{R}_+$, and apparently Zeman (1997) is not using the correct transformation (2) to obtain a density on $\mathbb{R}^2$.

We try another approach where we fit a non-parametric kernel density estimate $\hat{f}$ to the distances between home and place of infection in the LB data. The density $\hat{f}$ is subsequently transformed by (2) into a density $\hat{g}$. The kernel estimate of the population intensity is finally

$$\hat{\rho}(s) = \lambda \sum_{j \in U} K_j \hat{g}(\|s - u_j\|) \tag{3}$$

where U is the index set of the administrative units, K_j is the number of persons associated with the jth unit, and u_j is the census point of the unit, cf. Figure 3. Here λ is for the moment left unspecified as it is absorbed into another parameter introduced in Section 3.3. Note that we are ignoring the fact that people in the jth unit live in the vicinity of u_j and not exactly at u_j.

The kernel estimate and alternative models for the population intensity are further discussed in Section 4.

3.3 Prior distributions and likelihood using a LGCP

We model π in (1) by a log linear model,

$$\pi(s) = \exp\left(\beta^{\mathsf{T}} d(s) + Y(s)\right) \tag{4}$$

where $Y(s)$ is a zero-mean Gaussian process, $\beta = (\beta_0, \ldots, \beta_6)^{\mathsf{T}}$ is a regression parameter, and $d(s) = (1, d_1(s), \ldots, d_6(s))^{\mathsf{T}}$ where $d_1(s), \ldots, d_6(s)$ are the 6 covariates associated with the location $s \in S$, cf. Section 2.1.

The role of $\exp(Y(s))$ is partly to model deviations of $\rho(s)/\hat{\rho}(s)$ from one. Therefore we do not constrain (4) to be less than one. Actually, in Section 4, λ is absorbed in $\exp(\beta_0)$ and we replace the unknown ρ by the estimate (3) with $\lambda = 1$. Then $\pi(s)$ is more precisely a *relative risk function*, since for $s_1, s_2 \in S$, $\pi(s_1)/\pi(s_2)$ is the ratio of risk functions at the locations s_1 and s_2.

We assume that Y is second-order stationary and isotropic with exponential covariance function, i.e.

$$\mathrm{Cov}(Y(s_1), Y(s_2)) = c(\|s_1 - s_2\|; \sigma^2, \alpha) = \sigma^2 \exp(-\|s_1 - s_2\|/\alpha) \tag{5}$$

where $\sigma^2 > 0$ is the variance and $\alpha > 0$ is the correlation parameter. A log Gaussian Cox process (LGCP) is then obtained by assuming that conditionally on $Y = (Y(s))_{s \in S}$ and $\theta = (\beta, \sigma, \alpha)$, the cases X form a Poisson process with intensity function $\hat{\rho}(s)\pi(s)$.

We view the Gaussian distribution for Y as a prior and the conditional distribution of X given (Y, θ) as the likelihood. Furthermore, a hyper prior density $p(\theta)$ for θ is imposed; specific hyper priors are considered in Section 4. Notice that the likelihood depends on θ only through β, and it has density

$$p(x|Y, \beta) = \exp\left(|S| - \int_S \hat{\rho}(s) \exp\left(\beta^{\mathsf{T}} d(s) + Y(s)\right) \mathrm{d}s\right) \times \prod_{\xi \in x} \hat{\rho}(\xi) \exp\left(\beta^{\mathsf{T}} d(\xi) + Y(\xi)\right) \tag{6}$$

with respect to the unit rate Poisson process on S where $|\cdot|$ denotes area.

3.4 Posterior

The posterior, that is, the conditional distribution of (Y, θ) given $X = x$, can be specified as follows. Suppose that $p(\theta)$ is proper and let E_θ denote expectation conditionally on θ. For $n \geq 1$ and pairwise distinct $s_1, \ldots, s_n \in S$, let $f_{(s_1,\ldots,s_n)}(\cdot|\theta)$ denote the conditional density of $(Y(s_1), \ldots, Y(s_n))$ given θ. The posterior density of $(Y(s_1), \ldots, Y(s_n), \theta)$ given $X = x$ is defined by

$$\begin{aligned} f_{(s_1,\ldots,s_n)}(y_1, \ldots, y_n, \theta|x) \propto E_\theta[p(x|Y, \beta)|Y(s_1) = y_1, \ldots, Y(s_n) = y_n] \times \\ f_{(s_1,\ldots,s_n)}(y_1, \ldots, y_n|\theta)p(\theta). \end{aligned} \tag{7}$$

The posterior of the process (Y, θ) given $X = x$ is then given by the consistent set of finite-dimensional posterior distributions with densities of the form (7) for $n \geq 1$ and distinct $s_1, \ldots, s_n \in S$. If $p(\theta)$ is improper we define the posterior similarly provided it is well-defined; i.e. provided

$$\int E_\theta[p(x|Y, \beta)]p(\theta)\mathrm{d}\theta < \infty.$$

3.5 Discretized LGCPs

The integral in (6) depends on the infinitely dimensional random field Y which cannot be represented on a computer. In practice we therefore approximate the integral by a Riemann sum as follows. The region S is appropriately translated and embedded in a rectangular region, say a square $[0, b[^2$ of sidelength $b > 0$. For integer $M \geq 1$, define the lattice

$$I^M = \{b/(2M), b/M + b/(2M), \ldots, b(M-1)/M + b/(2M)\}^2$$

and let

$$C^M_{(u,v)} = S \cap [u - b/(2M), u + b/(2M)[\times[v - b/(2M), v + b/(2M)[, \quad (u, v) \in I^M.$$

The approximation of the integral is now

$$\int_S \hat{\rho}(s) \exp\left(\beta^\mathsf{T} d(s) + Y(s)\right)\mathrm{d}s \approx \sum_{\eta \in I^M} |C^M_\eta| \hat{\rho}(\eta) \exp\left(\beta^\mathsf{T} d(\eta) + Y(\eta)\right). \tag{8}$$

For $\xi \in x \cap C^M_\eta, \eta \in I^M$, it is for computational reasons also convenient (see Section 3.6) to approximate $\hat{\rho}(\xi)\exp\left(\beta^\mathsf{T} d(\xi) + Y(\xi)\right)$ by $\hat{\rho}(\eta)\exp\left(\beta^\mathsf{T} d(\eta) + Y(\eta)\right)$. For $\eta \in I^M$, we further replace $d(\eta)$ by $\tilde{d}^M(\eta) = \int_{C^M_\eta} d(s)\mathrm{d}s/|C^M_\eta|$, since the average value $\tilde{d}^M(\eta)$ is a better representative of the covariate values $d(s)$, $s \in C^M_\eta$, than $d(\eta)$. Combining this with (8), an approximation of $p(x|Y, \beta)$ is obtained:

$$p^M(x|Y,\beta) = \exp\Big(|S| - \sum_{\eta\in I^M} |C_\eta^M|\hat\rho(\eta)\exp\big(\beta^\mathsf{T}\tilde d^M(\eta) + Y(\eta)\big)\Big)\times \prod_{\substack{\eta\in I^M:\\ C_\eta^M\cap x\neq\emptyset}} \hat\rho(\eta)^{n^M(\eta)}\exp\Big(n^M(\eta)\big(\beta^\mathsf{T}\tilde d^M(\eta)+Y(\eta)\big)\Big) \quad (9)$$

where $n^M(\eta) = \mathrm{card}(x\cap C_\eta^M)$, $\eta\in I^M$. The approximate posterior density is hence given by

$$f^M_{(s_1,\ldots,s_n)}(y_1,\ldots,y_n,\theta|x) \propto E_\theta[p^M(x|Y,\beta)|Y(s_1)=y_1,\ldots,Y(s_n)=y_n]\times f_{(s_1,\ldots,s_n)}(y_1,\ldots,y_n|\theta)p(\theta). \quad (10)$$

It is defined similarly for $p(\theta)$ improper provided

$$\int E_\theta[p^M(x|Y,\beta)]p(\theta)\mathrm{d}\theta < \infty.$$

Our Bayesian approach is based on (9) where $\{s_1,\ldots,s_n\}$ agrees with I^M. Then we are in a sense back in an area level approach, since $p^M(x|Y,\beta)$ only depends on x through the counts $n^M(\eta)$, $\eta\in I^M$. However, we use a much finer partitioning of S into cells C_η^M than in typical area level approaches, see Section 4. Furthermore, under certain conditions we verify in the Appendix that expectations computed with respect to the approximate posterior (10) converge to the corresponding expectations with respect to (7) when M tends to infinity.

3.6 Markov chain Monte Carlo for discretized LGCPs

In this section we discuss MCMC simulation of the approximate posterior (10) when $\{s_1,\ldots,s_n\}$ agrees with I^M. The main obstacle is to handle the high dimensional covariance matrix of $\tilde Y = (Y(\eta))_{\eta\in I^M}$. However, the computational cost can be reduced very much by employing the circulant embedding technique described in Dietrich and Newsam (1993) and Wood and Chan (1994); see also Møller *et al.* (1998). This is shortly explained below.

The lattice I^M is extended to

$$I^M_{\mathrm{ext}} = \{b/(2M), b/M + b/(2M),\ldots, b(M_{\mathrm{ext}}-1)/M + b/(2M)\}^2$$

where $M_{\mathrm{ext}} \geq 2M-2$. For $\xi,\eta\in I^M_{\mathrm{ext}}$ let $\mathrm{gd}(\xi,\eta)$ denote geodesic distance when I^M_{ext} is wrapped on a torus. The symmetric matrix $K = \{c(\mathrm{gd}(\xi,\eta);\sigma^2,\alpha)\}_{\xi,\eta\in I^M_{\mathrm{ext}}}$ is block circulant with M_{ext} circulant blocks of dimension $M_{\mathrm{ext}}\times M_{\mathrm{ext}}$, so it can easily be diagonalized by means of the two-dimensional discrete Fourier transform. Note that the submatrix $\{c(\mathrm{gd}(\xi,\eta);\sigma^2,\alpha)\}_{\xi,\eta\in I^M}$ coincides with the covariance matrix of $\tilde Y$.

Suppose that K is positive semi-definite (i.e. K has non-negative eigenvalues). Then we can extend $\tilde{Y}$ to a larger Gaussian field $\tilde{Y}_{\text{ext}} = (\tilde{Y}_\eta)_{\eta \in I^M_{\text{ext}}}$ with covariance matrix K. Using a lexicographic ordering of the indices, we can set

$$\tilde{Y}_{\text{ext}} = Q^\mathsf{T}\Gamma \tag{11}$$

where Γ follows a standard multivariate normal distribution of dimension $d = \text{rank}(K)$, and where Q is a $M^2_{\text{ext}} \times d$ matrix of rank d so that $K = Q^\mathsf{T}Q$. Thus from Γ we obtain $\tilde{Y}_{\text{ext}}$ and hence $\tilde{Y}$.

Note that $M_{\text{ext}} = 2M-2$ is the smallest extension which gives a block circulant matrix K. If K turns out not to be positive semi-definite, it may help to use a larger extension $M_{\text{ext}} > 2M-2$. Such large extensions are usually only needed if the correlations for Y are slowly decaying. In order to apply the fast Fourier transform for the calculations of Q and $\tilde{Y}_{\text{ext}}$ by (11), M_{ext} should be a power of two (or three or five).

The posterior density for (Γ, θ) is

$$\pi(\gamma, \theta | x) \propto p(\theta) \exp(-\|\gamma\|^2/2) \exp\Big(- \sum_{\eta \in I^M} |C^M_\eta| \hat{\rho}(\eta) \exp\big(\beta^\mathsf{T} \tilde{d}^M(\eta) + \tilde{y}_\eta\big)\Big) \times$$
$$\prod_{\substack{\eta \in I^M: \\ C^M_\eta \cap x \neq \emptyset}} \hat{\rho}(\eta)^{n^M(\eta)} \exp\big(n^M(\eta)(\beta^\mathsf{T} \tilde{d}^M(\eta) + \tilde{y}_\eta)\big)$$

where $\tilde{y} = Q^\mathsf{T}\gamma$. This is very similar to the posterior density for the Poisson-log normal model considered in Christensen and Waagepetersen (2001). For the MCMC simulations of (Γ, θ) given $X = x$ we use the same type of fixed-scan hybrid algorithm as in Christensen and Waagepetersen (2001), where γ, β, σ, and α are updated in turn using so-called truncated Langevin-Hastings updates for γ and β and standard random walk Metropolis updates for σ and α. For further details of our MCMC algorithm we refer to the above-mentioned paper. Finally, posterior simulations of (Γ, θ) are transformed into posterior simulations of $(\tilde{Y}, \theta)$.

For fixed values of θ the nice convergence properties of truncated Langevin-Hastings algorithms applied to conditional simulation in discretized LGCPs, Poisson-log normal and related models are studied in Møller *et al.* (1998) and Christensen *et al.* (2001).

4 Results and discussion

In this section we discuss the results for the TBE data obtained by the Bayesian approach described above. Details concerning specification of priors, estimation of population intensity, and the MCMC algorithm are given in Section 4.1. Section 4.2 compares posterior results obtained with different models for the population intensity. Model selection is adressed in Section 4.3. The estimated relative risk map is discussed in Section 4.4. Concluding remarks are given in Section 4.6.

4.1 Specification of models and computational details

For the discretized LGCP, S is rescaled and embedded in a unit square which is divided into a grid of square cells C_η^M, $\eta \in I^M$ where $M = 65$, cf. Section 3.5. Thereby Central Bohemia is covered by 2166 cells, and S occupies about 51% of the unit square. Results with other values $M = 17,\ 33$, and 129, are discussed in Section 4.5.

The following models of the population intensity are considered, setting $\lambda = 1$, cf. Section 3.3.

- Model W (constant ρ): $\hat{\rho}_{\text{const}}(s) = \sum_{j\in U} K_j/|S|$ is constant.
- Model P (kernel based on paired data): $\hat{\rho}_{\text{pair}}(s) = \sum_{j\in U} K_j \hat{g}(\|s - u_j\|)$ is estimated as in (3) using the paired LB data.
- Models B,D, and E (Gaussian kernel): $\hat{\rho}_\tau(s) = \sum_{j\in U} K_j g(\|s - u_j\|; \tau)$ where $g(\cdot;\tau) : \mathbb{R}_+ \to \mathbb{R}_+$ is given by $g(h;\tau) = \phi(s;\tau)$ for $s \in \mathbb{R}^2$ with $\|s\| = h$ and where ϕ is the density for a two-dimensional radially symmetric Gaussian distribution with zero mean and standard deviation τ. Model B: $\tau = 0.7\,\text{km}$. D: $\tau = 2.5\,\text{km}$. E: $\tau = 5\,\text{km}$.

For the corresponding LGCPs, model W is equivalent to the limiting case with a Gaussian kernel where $\tau \to \infty$. Figure 4 shows the kernels used to compute population intensity models P, B, D, and E. Note that the tail for P falls between the tails of D and E.

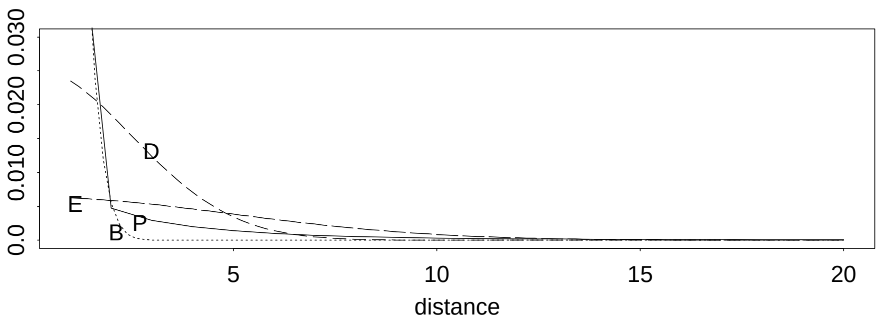

Figure 4: Kernels for models P, B, D, and E.

For all the population intensity models we use independent hyper priors for β, σ, and α given by

$$p_1(\beta) \propto 1, \quad \theta \in \mathbb{R}^7,$$

$$p_2(\sigma) \propto \exp(-10^{-6}/\sigma)/\sigma, \quad \sigma > 0,$$
$$p_3(\alpha) \propto 1/\alpha, \quad -6.91 < \log\alpha < -1.10.$$

The improper prior p_1 is completely flat and the improper p_2 yields an essentially flat prior for $\log\sigma$. The limits for the log uniform prior p_3 are chosen subjectively in order to accomodate a reasonable range of strengths of correlation. By similar arguments as in the proof of Proposition 1 in Christensen *et al.* (2001), a proper posterior (10) is obtained for the discretized LGCP, but strictly speaking we do not know whether a proper posterior (7) is also obtained for the original LGCP. One may therefore consider the possibility of restricting the supports of β and σ to large but bounded regions, see also the Appendix.

We use the hybrid algorithm mentioned at the end of Section 3.6 for the approximate posterior computations. In order to improve the mixing properties of the algorithm we use a reparametrization where each of the covariates $d_1, \ldots, d_6$ is standardized by substracting the mean and dividing by the standard deviation of the variable. However, when presenting our results in Section 4.2 we use the original parametrization. User-specified parameters of the algorithm are tuned to obtain acceptance rates which are fairly close to theoretically optimal values (see Christensen and Waagepetersen, 2001): the acceptance rates for updates of γ, β, σ, and $\log\alpha$ are 0.56, 0.54, 0.24, and 0.25, respectively. For posterior simulations we use 500,000 scans of the algorithm, where a scan consists of an update of each of the parameters γ, β, σ, and α. The Monte Carlo estimates of the various characteristics of the posterior shown in Section 4.2 are calculated from a subsample obtained with spacing equal to 100.

4.2 Posterior results and comparison of models

Posterior histograms of $\beta_0, \ldots, \beta_6$, σ, and $\log\alpha$ are rather symmetric and close to the normal distribution under any of the population intensity models; as an example the results for model B are shown in Figure 5. Note that the marginal posterior distribution for α is concentrated on a much smaller interval than the support of its prior p_3.

Posterior means for the β_i, and for the different models are shown in Table 1. The numbers in parantheses are the probabilities $p_i = P(\beta_i > 0|x)$. Under model B, p_4 and p_5 indicate that the presence of mixed forest (β_4) or foliate forest (β_5) increases the risk of infection; in Jiruše *et al.* (2000) the presence of mixed forest is concluded to be a significant covariate. For all the population intensity models there is evidence that the presence of coniferous forest decreases the risk of infection and, except for B, that a high altitude increases the risk of infection. The posterior means are rather sensitive to the choice of population intensity model. The qualitative results based on the posterior probabilities p_i are on the other hand rather similar for all population intensity models except model B.

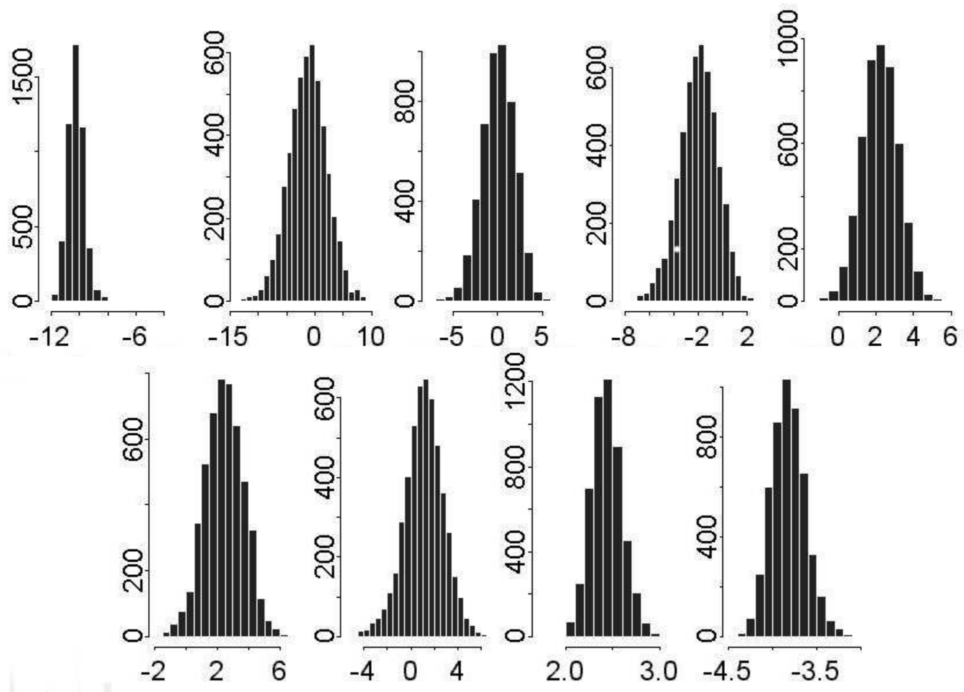

Figure 5: From left top to right bottom: marginal posterior distributions for $\beta_0, \ldots, \beta_6$, σ, and $\log \alpha$ under the model B.

The posterior means and standard deviations for σ and $\log \alpha$ are rather comparable for the different models. The posterior means for σ ranges between 2.0 (model W) and 2.4 (B), and the standard deviation between 0.1 (W) and 0.2 (model D). The posterior means for $\log \alpha$ are between -3.8 (B) and -3.3 (D), while the standard deviations take the value 0.2. The posterior mean of the empirical mean of $\tilde{Y}$ is close to zero for all models, and the posterior mean of its empirical standard deviation is a bit larger than 2 and close to the posterior mean of σ for all models. Notice that $\tilde{Y}$ is playing an important role in the model since the posterior for σ is concentrated on an interval far from zero. For the exponential correlation function with $\log \alpha = -3.8$, the correlation is bigger than 0.01 for distances less than 15 km on the physical scale. Finally, let

$$\Lambda^M(s) = \hat{\rho}(\eta) \exp\left(\beta^\mathsf{T} \tilde{d}^M(\eta) + Y(\eta)\right), \; s \in C_\eta^M, \tag{12}$$

denote the intensity function of the discretized LGCP. The posterior mean $E[\int_S \Lambda^M(s) \mathrm{d}s \,|\, x]$ of the intensity function integrated over Central Bohemia is between 445.5 and 446.1 (close to the number of observed cases) for the different models.

	β_0	β_1	β_2	β_3	β_4	β_5	β_6
W	-10.1 (.0)	-1.2 (.3)	-.5 (.4)	-4.5 (.0)	.1 (.6)	.2 (.6)	1.5 (1.0)
P	-9.7 (.0)	-1.0 (.4)	-.1 (.5)	-3.3 (.0)	.2 (.6)	.4 (.7)	1.0 (1.0)
B	-9.8 (.0)	-1.6 (.3)	-.0 (.5)	-1.8 (.1)	2.2 (1.0)	2.2 (1.0)	-0.1 (.5)
D	-9.6 (.0)	-2.1 (.2)	-.2 (.5)	-3.4 (.0)	-.2 (.4)	.3 (.6)	1.0 (1.0)
E	-9.8 (.0)	-1.6 (.3)	-.4 (.4)	-4.0 (.0)	-.3 (.3)	.0 (.5)	1.3 (1.0)

Table 1: Posterior means for β_i and $p_i = P(\beta_i > 0 \,|\, x)$ (in paranthesis), $i = 0, \ldots, 6$, under models W, P, B, D, and E.

4.3 Model selection

As the posterior results depend much on the choice of population map, one may naturally ask from which model conclusions should be drawn. In the Bayesian framework there exist several tools for model selection including Bayes factors, posterior predictive distributions, and, of course, an extended Bayesian analysis where prior probabilities are also assigned to the different models in question.

We restrict attention to the consideration of posterior predictive distributions, basically because this is supported by our present software. Consider a summary statistic $U(x)$ computed from the data x. The idea is to assess the fit of a posterior model by comparing $U(x)$ with the posterior predictive distribution; i.e. in our case the distribution of $U(X)$ where X is a Cox process with random intensity surface distributed as $[\Lambda^M|x]$, see (12). Below we consider two types of summary statistics: the counts $n^M(\eta)$, $\eta \in I^M$, and a variant of the K-function.

For the counts $n^M(\eta)$, $\eta \in I^M$, we just compare $n^M(\eta)$ with the posterior predictive mean $\hat{\lambda}_\eta = |C^M_\eta| E[\Lambda^M(\eta)|x]$ and compute discrepancy statistics

$$\chi^2_1 = \sum_{\eta \in I^M \cap S} (n^M(\eta) - \hat{\lambda}_\eta)^2 \quad \text{and} \quad \chi^2_2 = \sum_{\eta \in I^M \cap S} (n^M(\eta) - \hat{\lambda}_\eta)^2 / \hat{\lambda}_\eta.$$

The values of these χ^2-statistics can be used to rank the different models according to their predictive performance. Note that χ^2_2 is more tolerant towards deviations between $n^M(\eta)$ and the posterior predictive mean $\hat{\lambda}_\eta$ when $\hat{\lambda}_\eta$ is large. The values of χ^2_1 under the different models are B: 120, W: 171, E: 179, P: 184, D: 192. However, the picture is different for χ^2_2 where we have W: 1054, P: 1096, E: 1112, D: 1133, B: 1344. This is consistent with the degree of smoothness of the population maps employed so that the smallest values of χ^2_2 are obtained for the models with the smoothest population maps. Another approach for using the cell counts $n^M(\eta)$ would be to consider so-called cross-validation predictive densities (Gelfand, 1996), but this is computationally quite demanding in our setting. In the following we restrict attention to models B and W.

Our LGCP X can be extended to a so-called second-order intensity-reweighted stationary point process on $\mathbb{R}^2$ for which an extension of the K-function can be defined; for details see Baddeley *et al.* (2000). If $\lambda(\cdot)$ denotes the intensity function for X, the inhomogeneous K-function is given by

$$K_{\text{inhom}}(t) = \frac{1}{|A|} E\Big[\sum_{\xi \in X \cap A} \sum_{\substack{\eta \in X: \\ \xi \neq \eta}} \frac{1(\|\xi - \eta\| < t)}{\lambda(\xi)\lambda(\eta)} \Big] \tag{13}$$

for $t > 0$ and an arbitrary $A \subset \mathbb{R}^2$ with $0 < |A| < \infty$. It is common practice to transform K_{inhom} into $L_{\text{inhom}}(t) = \sqrt{K_{\text{inhom}}/\pi}$ which is equal to t for a Poisson process. From (13) we obtain an estimate of K_{inhom} by omitting the expectation, letting $A = S$, and replacing X with the observed data x; we here ignore the edge effects caused by unobserved tick infections outside S. Furthermore, the unknown $\lambda(\cdot)$ is replaced by the maximum likelihood estimate under the Poisson model corresponding to model W without Y (or equivalently with $\sigma^2 = 0$).

Figure 6 shows the estimated $L_{\text{inhom}}(t) - t$. Notice that the estimate is bigger than zero which indicates clustering — this is in accordance with the results in Section 4.2 which showed that Y was not a negligible part of the model. Theoretically, $L_{\text{inhom}}(0) = 0$, and the behaviour of the estimate for small values of t is an artifact due to the multiple points in the data. The dotted curves in Figure 6 are envelopes, i.e. pointwise minima and maxima for estimates of $L_{\text{inhom}}(t) - t$ computed from 39 point patterns simulated under the posterior predictive distributions corresponding to model B and W, respectively. If the observed data were generated by one of the posterior predictive distributions, then for each $t > 0$, there is 5% probability that the estimate of $L_{\text{inhom}}(t) - t$ from the data falls outside the envelopes. If we disregard the small t-values, then neither of the posterior predictive distributions seem to be in conflict with the observed data.

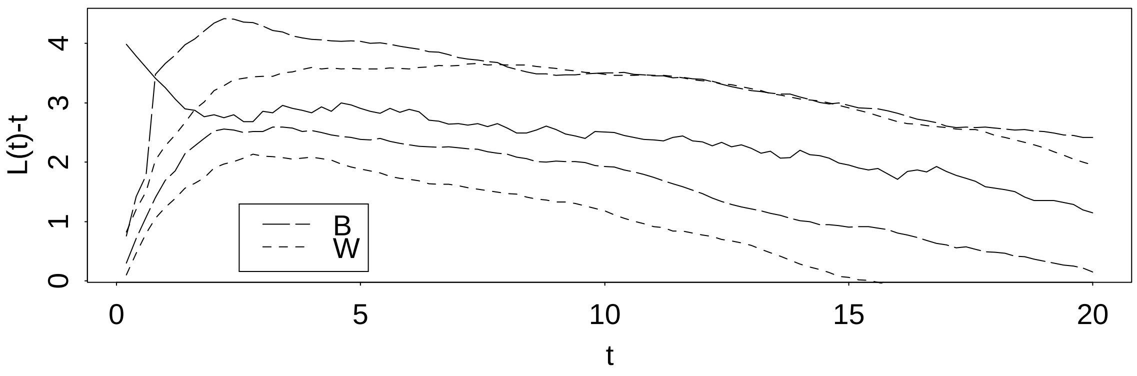

Figure 6: Estimated $L_{\text{inhom}}(t) - t$ (solid line) versus t (measured in km) under model B and model W.

4.4 Relative risk map

The posterior mean of the relative risk function under models B and W are shown in Figure 7 on a log scale. More precisely, in order to compare the results for model W and B we plot

$$\log E\big[\exp(\beta^T \tilde{d}^M(\eta) + Y(\eta) - \max_{\xi \in I^M \cap S} \{\beta^T \tilde{d}^M(\xi) + Y(\xi)\})|x\big], \eta \in I^M \cap S,$$

for each model. The relative risk function is less varying under model W than under model B; for model B the smallest and the largest values are $\exp(-11.06)$ and $\exp(-0.13)$, respectively, and for the model W the values are $\exp(-9.18)$ and $\exp(-0.05)$. Comparing Figure 7 with the maps in Zeman (1997), Mašata (1999), and Krejčíř (2000), the overall features are rather similar (no such map is shown in Jiruše *et al.*, 2000).

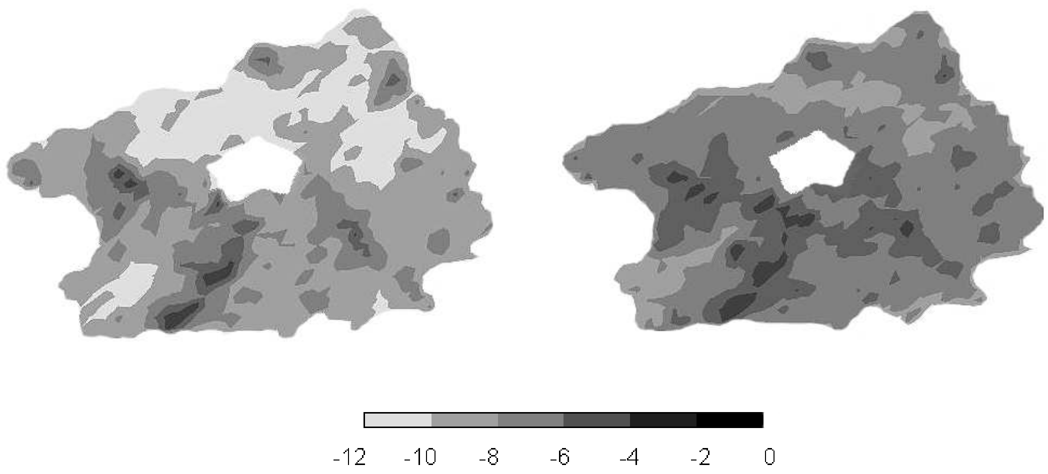

Figure 7: Maps of the logarithm of the posterior mean of the relative risk function divided by its maximal value. Left: model B. Right: model W.

Jiruše *et al.* (2000) compare their results with those of Mašata (1999) in a plot showing the credibility intervals of the relative risk function evaluated separately for each cell (ordered with increasing risk) in the irregular division of 41 cells used in Mašata (1999). Figure 8 shows 2.5% and 97.5% posterior quantiles for the log relative risk function. The uncertainty is large; for model B and the cell with the largest mean posterior relative risk the 2.5% and 97.5% quantiles are -1.32 and 0, respectively; for model W the corresponding numbers are -0.48 and -0.01, respectively. The oscillations of the quantiles are smaller for model W than model B due to the constant population intensity for model W.

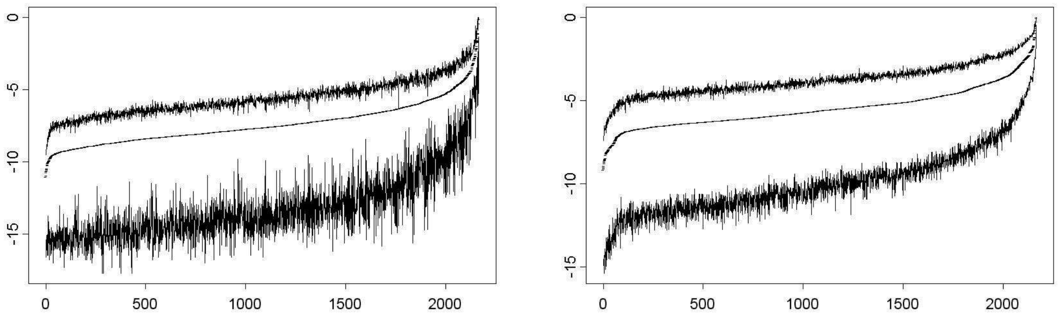

Figure 8: The log mean posterior relative risk function divided by its maximal value for the 200 cells with largest mean posterior relative risk. The lower and upper curves are 2.5% and 97.5% posterior quantiles. Left: model B. Right: model W.

4.5 Dependence on grid size

Tables 2 and 3 contain posterior means for β_i and $p_i = P(\beta_i > 0 \,|\, x)$, $i = 0, \ldots, 6$, obtained with different choices of the grid size M under model B and W, respectively. The numbers depend much on M, and in particular the different results for $M = 65$ and $M = 129$ suggest that we need a larger value than $M = 65$ in order to approximate well the exact posterior of the β_i's. However, using $M = 129$ or larger is computationally very time consuming. It is perhaps not so surprising that the results for $M = 65$ and $M = 129$ differ since the various covariates vary on a fine scale compared with the spatial resolution of the grid cells obtained with $M = 65$ and $M = 129$. The cell-averaged covariate vectors $\tilde{d}^{65}(\cdot)$ and $\tilde{d}^{129}(\cdot)$ (see Section 3.5) may thereby be quite different from each other and the original covariate vector $d(\cdot)$. Comparing model B and W, the results for $M = 65$ and $M = 129$ are more similar for model W since the constant population intensity is not sensitive towards averaging over grid cells.

M	β_0	β_1	β_2	β_3	β_4	β_5	β_6
17	-8.3 (.0)	-8.3 (.3)	-3.8 (.3)	-4.7 (.1)	6.8 (1.0)	4.9 (1.0)	-1.6 (.1)
33	-8.4 (.0)	-8.1 (.1)	-7.5 (.0)	0.3 (.6)	4.7 (1.0)	7.1 (1.0)	-2.6 (.0)
65	-9.8 (.0)	-1.6 (.3)	-.0 (.5)	-1.8 (.1)	2.2 (1.0)	2.2 (1.0)	-.1 (.5)
129	-11.4 (.0)	-1.2 (.2)	-1.2 (.1)	-5.0 (.0)	-.4 (.3)	.1 (.5)	2.0 (1.)

Table 2: Posterior means for β_i and $p_i = P(\beta_i > 0 \,|\, x)$ (in paranthesis), $i = 0, \ldots, 6$, under model B and for four choices of grid size $M = 17,\ 33,\ 65,\ 129$.

M	β_0	β_1	β_2	β_3	β_4	β_5	β_6
17	-9.4 (.0)	9.1 (.8)	-8.1 (.1)	-7.3 (.0)	6.9 (1.0)	3.3 (.9)	.8 (.7)
33	-9.6 (.0)	-1.7 (.4)	-4.0 (.1)	-3.4 (.0)	3.0 (1.0)	2.5 (1.0)	1.0 (.9)
65	-10.1 (.0)	-1.2 (.3)	-.5 (.4)	-4.5 (.0)	.1 (.6)	.2 (.6)	1.5 (1.0)
129	-11.3 (.0)	-.6 (.4)	-1.2 (.1)	-6.0 (.0)	-.8 (.1)	-.3 (.3)	2.1 (1.0)

Table 3: Posterior means for β_i and $p_i = P(\beta_i > 0 \,|\, x)$ (in paranthesis), $i = 0, \ldots, 6$, under model W and for four choices of grid size $M = 17,\ 33,\ 65,\ 129$.

4.6 Discussion

The approximate posterior results are very sensitive to the choice of grid size M for the discretization and results must thus be reported conditional on the chosen value of M. Considering $M = 65$, the results concerning which covariates are important for predicting tick infections depend much on which population model is used. The best fit is according to the statistic χ^2_2 obtained with the population intensity model W. A uniform population intensity is on the other hand not realistic. Model P is obtained empirically from paired data, but under very crude assumptions. With the data available it seems hard to make a definite choice between the different population intensity models considered. One should therefore consider either of the two following possibilities: 1) collect more data from which a satisfactory population intensity model could be constructed, or 2) include uncertainty concerning the population model in the analysis e.g. by introducing a prior for the various models, or perhaps just on τ if one restricts attention to the Gaussian smoothing kernels.

In many examples of disease mapping one fixed population map is regarded as the truth, but our results suggest that such an approach can easily be misleading. Moreover, our analysis appears to be more satisfactory than those in Zeman (1997) and Krejčíř (2000), since they do not include the covariate information and because of larger flexibility in our approach. Jiruše *et al.* (2000) and Mašata (1999) deal with covariates but in an area level approach. Thanks to the point process setting used in the present paper, we have provided a more detailed modelling and analysis of the spatial dependence (recall that Mašata (1999) uses only 41 irregular cells and Jiruše *et al.* (2000) only 141 cells of size 10×10 km^2, and it is only in Mašata (1999) that spatial dependence is incorporated).

The sensitivity to the choice of discretization may be reduced by omiting the approximation where the product in (6) over observed points is replaced by the product in (9) over grid cells. One then needs to deal with the joint density of $(Y(\eta))_{\eta \in I^M \cup x}$. By factorizing the density of $(Y(\eta))_{\eta \in I^M \cup x}$, methods based on the fast Fourier transform may still be used, but the programming and computational complexity increases.

It would also be interesting to analyze the tick data by a Cox process model where $\exp(Y(\cdot))$ in (4) is replaced by the random intensity function of a Poisson/gamma process (Wolpert and Ickstadt, 1998) or more generally, a shot-noise Cox process (Brix, 1999), see also Møller and Waagepetersen (2001). The calculation of the integral in (6) may still require numerical methods, but possibly much finer approximations may be feasible, so that the posterior results become less sensitive to the numerical approximation. Moreover, some care must be taken in order to handle boundary effects.

Acknowledgements: The authors wish to thank Dr. Petr Zeman, Regional Center of Hygiene, Dittrichova 17, Prague 2, Czech Republic, for encouraging their interest in the problem and for providing the data sets. VB was supported by grants GAČR 201/98/0090 and MSM 113200008 and KB acknowledges the support and GAČR 201/99/0269. JM and RW were supported by MaPhySto – Centre for Mathematical Physics and Stochastics, funded by a grant from The Danish National Research Foundation, and by the European Union's network "Statistical and Computational Methods for the Analysis of Spatial Data. ERB-FMRX-CT96-0096".

Appendix: Convergence of approximate posteriors

In our Bayesian analysis we are interested in expectations with respect to the posterior distribution of $(Y(s_1), \dots, Y(s_n), \theta)$ for a finite set of locations $s_1, \dots, s_n$, where in practice we compute expectations with respect to the approximate posterior (10). It is therefore natural to investigate whether approximate expectations converge to the corresponding exact posterior expectations with respect to (7) when M tends to infinity. In Theorem 1 below we verify that this is the case under the somewhat restrictive assumption that the hyper prior density $p(\theta)$ has compact support. Furthermore, we let $Y(s)_{s\in\mathbb{R}^2}$ denote a second-order stationary zero-mean Gaussian process with almost surely continuous sample paths and variance $\sigma^2 > 0$. These properties are satisfied for a Gaussian field with the exponential covariance function in (5), cf. Theorem 3.4.1 in Adler (1981).

The following lemma is used in the proof of Theorem 1.

Lemma 1. *For any $l > 0$, $\psi^2 > \sigma^2$, and compact $K \subset \mathbb{R}^2$ with positive area,*

$$E \exp\big(\sup_{s\in K} lY_s + (\sup_{s\in K} Y_s)^2/(2\psi^2)\big) < \infty.$$

Proof. Let $0 < \epsilon < (2\sigma^2)^{-1} - (2\psi^2)^{-1}$. Since

$$\lim_{t\to\infty} t^{-2} \log P(\sup_{s\in K} Y_s > t) = (2\sigma^2)^{-1} \tag{14}$$

(see e.g. p. 160 in Adler, 1981) there exists a $T > 0$ such that

$$P(\sup_{s\in K} Y_s > t) < \exp(t^2(\epsilon - (2\sigma^2)^{-1}))$$

for all $t > T$. Let $F(t) = \exp(lt + t^2/(2\psi^2))$, $G(t) = P(\sup_{s\in K} Y_s > t)$, and note that $G(0) = 1$. The result now follows by integration by parts:

$$E\exp\left(\sup_{s\in K} lY_s + (\sup_{s\in K} Y_s)^2/(2\psi^2)\right) = 1 + \int_0^\infty G(t)F(dt) <$$
$$\int_0^T G(t)F(dt) + \int_T^\infty \exp\big(t^2(\epsilon - (2\sigma^2)^{-1} + (2\psi^2)^{-1}) + lt\big)\big(l + t/(\psi^2)\big)\mathrm{d}t < \infty.$$

$\square$

Theorem 1. *Assume that the support* $\Theta_0 = \{\theta \in \mathbb{R}^8 : p(\theta) > 0\}$ *is compact,* $\Lambda(s) = \hat{\rho}(s)\exp\big(\beta^\mathsf{T} d(s) + Y(s)\big)$ *is almost surely Riemann integrable on* S*, and* $\hat{\rho}(\cdot)$ *and* $d(\cdot)$ *are bounded on* S *and continuous at all* $\eta \in x$*. Consider any Borel function* $h : \mathbb{R}^n \times \Theta_0 \to \mathbb{R}$ *with* $h(y_1, \ldots, y_n, \theta) \le \exp(\max_i y_i^2/(2\psi^2))k(\theta)$ *for all* $(y_1, \ldots, y_n, \theta)$*, where* $\sup_{\theta\in\Theta_0} \sigma^2 < \psi^2 < \infty$ *and* $k : \mathbb{R}^8 \to \mathbb{R}$ *is a Borel function so that* $\int k(\theta)p(\theta)\mathrm{d}\theta < \infty$*. Then*

$$\lim_{M\to\infty} \int h(y_1, \ldots, y_n, \theta) f^M_{(s_1,\ldots,s_n)}(y_1, \ldots, y_n, \theta|x)\mathrm{d}y\mathrm{d}\theta =$$
$$\int h(y_1, \ldots, y_n, \theta) f_{(s_1,\ldots,s_n)}(y_1, \ldots, y_n, \theta|x)\mathrm{d}y\mathrm{d}\theta.$$

Proof. Combining the fact that $(Y(s))_{s\in S}$ is almost surely continuous with the assumptions on $\hat{\rho}(\cdot)$, $d(\cdot)$, and $\Lambda(\cdot)$, we obtain that $\lim_{M\to\infty} p^M(x|Y,\beta) = p(x|Y,\beta)$ almost surely, cf. (6) and (9). Further, by (9)

$$p^M(x|Y,\beta) \le \sup_{s\in S} \hat{\rho}(s)^{\mathrm{card}(x)} \exp\Big(\mathrm{card}(x)\big(\sup_{s\in S} \beta^\mathsf{T} d(s) + \sup_{s\in S} Y(s)\big)\Big)$$

and so

$$h(Y(s_1), \ldots, Y(s_n), \theta)p^M(x|Y,\beta) \le$$
$$\sup_{s\in S} \hat{\rho}(s)^{\mathrm{card}(x)} \exp\Big(\mathrm{card}(x)\big(\sup_{s\in S} \beta^\mathsf{T} d(s)\big)\Big)k(\theta) \times$$
$$\exp\big(\mathrm{card}(x) \sup_{s\in S} Y(s) + \sup_{s\in S} Y(s)^2/(2\psi^2)\big).$$

By Lemma 1,

$$K(\theta) \equiv E_\theta \exp\big(\mathrm{card}(x) \sup_{s\in S} Y(s) + \sup_{s\in S} Y(s)^2/(2\psi^2)\big) < \infty. \tag{15}$$

Note that $K(\theta)$ depends only on (σ, α) and it can be checked that $K(\cdot)$ is continuous (we omit the details). Using (15) and the compactness of Θ_0 the result now follows by applying dominated convergence to the numerator and the denumerator in

$$\int h(y_1, \ldots, y_n, \theta) f^M_{(s_1,\ldots,s_n)}(y_1, \ldots, y_n, \theta|x) \mathrm{d}y \mathrm{d}\theta =$$
$$\int E_\theta[h(Y(s_1), \ldots, Y(s_n), \theta) p^M(x|Y, \beta)] p(\theta) \mathrm{d}\theta \Big/ E_\theta[p^M(x|Y, \beta)] p(\theta) \mathrm{d}\theta.$$

□

Remark: In applications (including ours) $\hat{\rho}(\cdot)$ and $d(\cdot)$ will typically be piecewise continuous, whereby the assumptions of Theorem 1 concerning $\hat{\rho}(\cdot)$ and $d(\cdot)$ are fulfilled. One is further typically interested in computing the posterior probability of some event (in which case $h(\cdot)$ is an indicator function) or moments of $Y(s_i)$ or $\pi(s_i)$, $i = 1, \ldots, n$. In all these cases the corresponding choices of $h(\cdot)$ are covered by the theorem.

References

Adler, R. (1981). *The Geometry of Random Fields*. Wiley.

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

Best, N. G., Arnold, R. A., Thomas, A., Waller, L. A. and Conlon, E. M. (1999). Bayesian models for spatially correlated disease and exposure data. In: *Bayesian Statistics 6* (eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), Oxford University Press, Oxford, 131–156.

Bithell, J. F. (1990). An application of density estimation to geographical epidemiology. *Statistics in Medicine* **9**, 691–701.

Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability* **31**, 929–953.

Christensen, O. F. and Waagepetersen, R. (2001). Bayesian prediction of spatial count data using generalised linear mixed models. *Biometrics* (to appear).

Christensen, O. F., Møller, J. and Waagepetersen, R. (2001). Geometric ergodicity of Metropolis Hastings algorithms for conditional simulation in generalised linear mixed models. *Methodology and Computing in Applied Probability* (to appear).

Coles, P. and Jones, B. (1991). A lognormal model for the cosmological mass distribution. *Monthly Notices of the Royal Astronomical Society* **248**, 1–13.

Dietrich, C. R. and Newsam, G. N. (1993). A fast and exact method for multidimensional Gaussian stochastic simulation. *Water Resources Research* **29**, 2861–2869.

Diggle, P. (2000). Overview of statistical methods for disease mapping and its relationship to cluster detection. In: *Spatial Epidemiology: Methods and Applications* (eds. P. Elliott, J. C. Wakefield, N. G. Best and D. J. Briggs), Oxford University Press, Oxford, 87–103.

Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns*. Academic Press, London.

Gelfand, A. E. (1996). Model determination using sampling-based methods. In: *Markov Chain Monte Carlo in Practice* (eds. W. R. Gilks, S. Richardson and D. J. Spiegelhalter), Chapman & Hall, London, 145–162.

Jiruše, M., Machek, J., Beneš, V. and Zeman, P. (2000). A Bayesian estimate of the risk of tick-borne diseases. *KPMS Preprint 08*, Charles University, Prague.

Knorr-Held, L. (2001). Some remarks on Gaussian Markov random field models for disease mapping. In: *Highly Structured Stochastic Systems* (eds. P. J. Green, N. L. Hjort and S. Richardson), Oxford University Press (to appear).

Krejčíř, P. (2000). A maximum likelihood estimator of an inhomogeneous Poisson point process intensity using beta-splines. *Kybernetika* **36**(4), 455–464.

Lawson, A. B., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.-F. and Bertollini, R., eds. (2001). *Disease Mapping and Risk Assessment for Public Health*. Wiley, Chichester.

Mašata, M. (1999). Assessment of risk of infection by means of a Bayesian method. In: *Proceedings of the International Conference on Stereoly, Spatial Statistics, and Stochastic Geometry* (eds. V. Beneš, J. Janáček and I. Saxl), Praha, 197–202.

Møller, J. (2001). A comparison of spatial point processes in epidemiological applications. In: *Highly Structured Stochastic Systems* (eds. P. J. Green, N. L. Hjort and S. Richardson), Oxford University Press (to appear).

Møller, J. and Waagepetersen, R. (2001). Statistical inference for Cox processes. In: *Spatial cluster modelling* (eds. D. Denison and A. B. Lawson), Chapman and Hall (to appear).

Møller, J., Syversveen, A.-R. and Waagepetersen, R. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics* **25**, 451–482.

Neyman, J. and Scott, E. L. (1958). Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society B* **20**, 1–43.

Richardson, S. (2001). Spatial models in epidemiological applications. In: *Highly Structured Stochastic Systems* (eds. P. J. Green, N. L. Hjort and S. Richardson), Oxford University Press (to appear).

Stern, H. S. and Cressie, N. (1999). Inference for extremes in disease mapping. In: *Disease Mapping and Risk Assessment for Public Health* (eds. A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J.-F. Viel and R. Bertollini), Wiley, New York, 63–84.

Wolpert, R. L. and Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**, 251–267.

Wood, A. and Chan, G. (1994). Simulation of stationary Gaussian processes in $[0,1]^d$. *Journal of Computational and Graphical Statistics* **3**, 409–432.

Zeman, P. (1997). Objective assessment of risk maps of tick-borne encephalitis and lyme boreliosis based on spatial patterns of located cases. *International Journal of Epidemiology* **26**, 1121–1130.