Biometrika (2010), **xx**, x, *pp*. 1–14

Printed in Great Britain

Bayesian geostatistical modeling with informative sampling locations

BY D. PATI

Department of Statistical Science, Box 90251, 214 Old Chemistry Building, Duke University, Durham, North Carolina 27708-0251, USA

dp55@stat.duke.edu

B. J. REICH

Department of Statistics, 2311 Stinson Drive, 4264 SAS Hall, Box 8203, North Carolina State University, Raleigh, North Carolina 27695, USA

brian_reich@ncsu.edu

D.B. DUNSON

Department of Statistical Science, Box 90251, 214 Old Chemistry Building, Duke University, Durham, North Carolina 27708-0251, USA

dunson@stat.duke.edu

SUMMARY

We consider geostatistical models that allow the locations at which data are collected to be informative about the outcomes. A Bayesian approach is proposed, which models the locations using a log Gaussian Cox process, while modeling the outcomes conditionally on the locations as Gaussian with a Gaussian process spatial random effect and adjustment for the location intensity process. We prove posterior propriety under an improper prior on the parameter controlling the degree of informative sampling, demonstrating that the data are informative. In addition, we show that the density of the locations and mean function of the outcome process can be estimated consistently under mild assumptions. The methods show significant evidence of informative sampling when applied to ozone data over the Eastern United States.

Some key words: Cox process; Gaussian process; Joint model; Point pattern; Posterior consistency; Preferential sampling.

1. INTRODUCTION

Geostatistical models focus on inferring a continuous spatial process based on data observed at finitely many locations, with the locations typically assumed to be noninformative. As noted by Diggle et al. (2010), this assumption is commonly violated for point-referenced spatial data, as it is not unusual to collect data at locations thought to have a large or small value for the outcome. For example, in monitoring of air pollution, one may place more monitors at locations believed to have a high value of ozone or another pollutant, while in studying distribution of animal species one may systematically look in locations thought to commonly contain the species of interest. Diggle et al. (2010) proposed a shared latent process model to adjust for bias due to informative sampling locations. Their analysis was implemented using a Monte Carlo approach for maximum likelihood estimation.

We follow a Bayesian approach using a model related to those described by Menezes (2005), Ho & Stoyan (2008) and Diggle et al. (2010). The locations are modeled using a log Gaussian Cox process (Møller et al., 2001), with the intensity function included as a spatially-varying predictor in the outcome model, which also includes spatial random effects drawn from a Gaussian process. A parameter a controls the degree of informative sampling, and the sampling locations are ignorable in the special case in which $a = 0$, while $a > 0$ implies a tendency to take more observations at spatial locations having relatively high outcome values. This model modifies shared random effects models for joint modeling of longitudinal and event time data (Radcliffe et al., 2004) and for accommodating informative missingness (Wu & Follmann, 1999).

To our knowledge, we are the first to develop a Bayesian approach to the informative locations problem in geostatistical modeling. However, adapting recently proposed models to the Bayesian paradigm is relatively straightforward, and our primary contribution is studying the theoretical properties of the model. In particular, it is not obvious that the data contain information about the informativeness of the sampling locations, and one may wonder to what extent the prior is driving the results even in large samples. We address this concern by proving that the posterior is proper under a noninformative prior on a. In addition, one can consistently estimate a, the density of the sampling locations and the mean function of the outcome process. This later result extends recent work showing posterior consistency in Gaussian process regression models (Choi & Schervish, 2007; Choi, 2007). Proofs are provided in an Appendix.

2. INFORMATIVE SAMPLING MODEL

Our objective is to estimate the spatial surface $\mu(s) \in \mathbb{R}$ for all $s \in \mathcal{D} \subset \mathbb{R}^2$ based on observations $y_1, \ldots, y_n$ at locations $s_1, \ldots, s_n \in \mathcal{D}$. We propose the following joint model

$$y_i \mid s_i \sim N\{\eta(s_i) + a\xi(s_i), \sigma^2\}, \quad p(s_i) = \frac{\exp\{\xi(s_i)\}}{\int_{\mathcal{D}} \exp\{\xi(s)\} ds} \quad (i = 1, \ldots, n), \tag{1}$$

where the observations are independent across locations s_i given $\xi(s)$ and $\eta(s)$, and $p(s)$ is the location density. Assuming the locations are a realization of an inhomogeneous Poisson process with log intensity $\xi(s)$, the mean surface is characterized as $\mu(s) = \eta(s) + a\xi(s)$, where $\eta(s)$ is a baseline surface and $a\xi(s)$ is an adjustment due to informative sampling. Letting $x(s)$ denote a vector of spatial covariates, $\xi(s) = x(s)^{\mathrm{T}}\beta_\xi + \xi_r(s)$ and $\eta(s) = x(s)^{\mathrm{T}}\beta_\eta + \eta_r(s)$, where β_ξ and β_η are regression coefficients and $\xi_r(s)$ and $\eta_r(s)$ are mean zero residual processes.

The log sampling density is treated as a latent covariate to adjust for informative sampling, with $a > 0$ implying that samples are more likely to be taken in areas with a large response. Setting the coefficient in β_ξ corresponding to the intercept to zero for identifiability,

$$E(y_i \mid s_i) = x(s_i)^{\mathrm{T}}\beta^* + a\xi_r(s_i) + \eta_r(s_i) \quad (i = 1, \ldots, n), \tag{2}$$

where $\beta^* = a\beta_\xi + \beta_\eta$. Therefore, accounting for informative sampling is only necessary when there is an association between the spatial surface of interest and the sampling density that cannot be explained by the shared spatial covariates $x(s)$.

The residuals $\xi_r(s) \sim \Pi_{\xi_r}$ and $\eta_r(s) \sim \Pi_{\eta_r}$ are assigned independent mean zero Gaussian process priors with Matérn covariance functions (Stein, 1999),

$$c(h \mid \psi) = \frac{\tau^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\nu^{1/2}h}{\rho}\right)^{\nu} \mathcal{K}_\nu \left(\frac{2\nu^{1/2}h}{\rho}\right), \quad h = ||s - s'||, \tag{3}$$

where $\psi = (\tau^2, \rho, \nu)$ and $\mathcal{K}$ is the modified Bessel function of the second kind. The Matérn covariance has three parameters: $\tau^2 > 0$ controls the variance, $\rho > 0$ controls the spatial range of the correlation, and $\nu > 0$ controls the smoothness of the process. Special cases include the exponential $c(h \mid \psi) = \tau^2 \exp(-2^{1/2}h/\rho)$ with $\nu = 1/2$, and the squared exponential $c(h \mid \psi) = \tau^2 \exp(-2h^2/\rho^2)$ with $\nu = \infty$.

3. THEORETICAL PROPERTIES

3·1. *Weak posterior consistency*

In this section we obtain posterior consistency of the parameters of our model with respect to fixed-domain asymptotics. Consider the joint model defined in §2, with $\mathcal{D} = [0, 1]^2$ without loss of generality and $\Pi_{\xi_r}, \Pi_{\eta_r}$ Gaussian processes on $\mathcal{C}(\mathcal{D})$, the space of continuous functions on $\mathcal{D}$. Letting $c(h \mid \psi_\xi)$ and $c(h \mid \psi_\eta)$ denote the covariance functions for ξ_r and η_r, respectively, we choose independent bounded hyperpriors for $\tau^2_\xi, \tau^2_\eta, \nu_\xi$ and ν_η while letting $\rho_\xi \sim \pi_\xi$ and $\rho_\eta \sim \pi_\eta$, where the supports of both π_η and π_ξ are $\mathbb{R}^+$. We choose a proper prior on $\mathbb{R}$ for a, $\beta_\xi \sim \mathrm{N}(\beta_{0\xi}, \Sigma_{0\xi})$, $\beta_\eta \sim \mathrm{N}(\beta_{0\eta}, \Sigma_{0\eta})$ and $\sigma^2 \sim$ Inv-Ga$(\alpha_\sigma, \beta_\sigma)$.

ASSUMPTION 1. *The prior $\zeta \sim \Pi$ satisfies the prior positivity condition $\Pi(\zeta : ||\zeta - \zeta_0||_\infty < \epsilon) > 0$ for all $\epsilon > 0$ and for any $\zeta_0 \in \mathcal{C}(\mathcal{D})$.*

van der Vaart & van Zanten (2009) showed that Assumption 1 holds for Gaussian process priors with squared exponential covariance under mild conditions, and Choi (2005) provided a set of sufficient conditions on the Matérn covariance kernel for the same setting.

ASSUMPTION 2. *The covariates are uniformly bounded, so there exists an $M > 0$ such that $||x(s)|| \leq M$ for all $s \in \mathcal{D}$.*

THEOREM 1. *Under model (1)–(2) with priors chosen as described in §3 and Assumptions 1–2, the posterior distribution $\Pi(\xi_r, \eta_r, a, \beta_\xi, \beta_\eta, \sigma \mid \{(y_i, s_i), i = 1, \ldots, n\})$, is weakly consistent.*

Theorem 1 does not imply that the hyperparameters in the covariance kernel are consistently estimated, though we do take into account uncertainty in these parameters and do not assume that the priors are well specified. It is typically not possible to consistently estimate all the parameters in the Matérn covariance (Zhang, 2004).

3·2. *Posterior propriety of a*

Under model (1)–(2), the parameter a controls the degree of informative sampling. The uniform improper prior, $\pi_a(a) \propto 1$, provides a noninformative choice. Theorem 2 shows that this prior leads to a proper posterior, implying that the data are informative about a.

Letting $s = (s_1, s_2, \ldots, s_n)$, $y = (y_1, y_2, \ldots, y_n)^{\mathrm{T}}$, $\xi^n_r = \{\xi_r(s_1), \xi_r(s_2), \ldots, \xi_r(s_n)\}^{\mathrm{T}}$ and $\eta^n_r = \{\eta_r(s_1), \eta_r(s_2), \ldots, \eta_r(s_n)\}^{\mathrm{T}}$, we have $\xi^n_r \sim \mathrm{N}(0, \Sigma^n_\xi)$ and $\eta^n_r \sim \mathrm{N}(0, \Sigma^n_\eta)$, where $\Sigma^n_\xi(s, s') = c(||s - s'|| \mid \psi_\xi)$ and $\Sigma^n_\eta(s, s') = c(||s - s'|| \mid \psi_\eta)$ for $s, s' \in \mathcal{D}$. Let $c(h \mid \psi_\xi) = \tau^2_\xi \exp(-2^{1/2}h^p/\rho_\xi)$ and $c(h \mid \psi_\eta) = \tau^2_\eta \exp(-2^{1/2}h^p/\rho_\eta)$ for $0 < p \leq 2$. We assume independent bounded priors on τ_ξ and τ_η and independent discrete uniform priors on ρ_ξ and ρ_η. Also, $\beta_\xi \sim \mathrm{N}(\beta_{0\xi}, \Sigma_{0\xi})$, $\beta_\eta \sim \mathrm{N}(\beta_{0\eta}, \Sigma_{0\eta})$ and $\sigma^2 \sim \pi(\sigma^2)$. Here we focus on powered exponential covariance functions rather than Matérn to simplify calculations. A similar result should hold for Matérn covariance functions if the priors on the hyperparameters have a bounded support.

THEOREM 2. *With the above prior specifications, the marginal posterior distribution of a, $p(a \mid y, s)$, is proper provided $n \geq 2$ and $E_\pi(\sigma) < \infty$.*

When the conditions of Theorem 2 are satisfied, the joint posterior is also proper.

4. COMPUTATIONAL DETAILS

The exact density for the sample locations in (1) is not available analytically, so approximation is required. In point process modeling the integral is often approximated as the sum over a fine grid. Letting $t_1, \ldots, t_M \in \mathcal{D}$ be a rectangular grid covering $\mathcal{D}$ with cell area Δ,

$$\int_{\mathcal{D}} \exp\{\xi(s)\} ds \approx \Delta \sum_{j=1}^{M} \exp\{\xi(t_j)\}. \quad (4)$$

This approximation yields a tractable posterior, but requires computationally expensive matrix inversions, which we limit using a kernel convolution approximation to f.

Let $\delta(s)$ be a mean zero Gaussian process with covariance $c(h \mid \psi)$. A process convolution (Higdon, 2002) lets

$$\delta(s) = \int_{\mathcal{D}} K_\psi(s-u) dW(u), \quad (5)$$

where W is Brownian motion and K_ψ is a kernel with parameters ψ. The kernel corresponding to the Matérn covariance is

$$K_\psi(u) = \tau \frac{\Gamma(\nu+1)^{1/2} \nu^{\nu/4+1/4} |u|^{\nu/2-1/2}}{\pi^{1/2} \Gamma(\nu/2+1/2) \Gamma(\nu)^{1/2} \rho^{\nu/2+1/2}} \mathcal{K}_{\nu/2+1/2} \left(\frac{2\nu^{1/2}|u|}{\rho} \right).$$

The kernel convolution representation of the Gaussian process in (5) is often used to motivate dimension reduction for the spatial process. Let $\phi_1, \ldots, \phi_N$ be a grid of spatial knots. Then for large N

$$\delta(s) \approx \sum_{j=1}^{N} K_\psi(s-\phi_j) w_j, \quad (6)$$

where $w_j \sim \mathrm{N}(0,1)$. Applying kernel convolution to $\xi(s)$ and $\eta(s)$ yields

$$y_i \mid s_i \sim N\left\{ x(s_i)^{\mathrm{T}} \beta^* + \sum_{j=1}^{N} K_{\psi_\eta}(s_i - \phi_j) u_j + a \sum_{j=1}^{N} K_{\psi_\xi}(s_i - \phi_j) v_j, \sigma^2 \right\}, \quad (7)$$

$$p(s_i) = \frac{\exp\left\{ x(s_i)^{\mathrm{T}} \beta_\xi + \sum_{j=1}^{N} K_{\psi_\xi}(s_i - \phi_j) v_j \right\}}{\sum_{l=1}^{M} \exp\left\{ x(t_l)^{\mathrm{T}} \beta_\xi + \sum_{j=1}^{N} K_{\psi_\xi}(t_l - \phi_j) v_j \right\}},$$

where $u_j, v_j \sim \mathrm{N}(0,1)$. Selecting the number of grid points M and knots N is discussed in §5 & 6.

We use a combination of Gibbs and Metropolis sampling for posterior computation. Assuming conjugate normal and inverse gamma priors, and reparameterization so that $u_j \sim \mathrm{N}(0, \tau_\eta^2)$ and $v_j \sim \mathrm{N}(0, \tau_\xi{}^2)$, the full conditionals for β^*, a, τ_η^2, τ_ξ^2 and the vector $(u_1, \ldots, u_N)^{\mathrm{T}}$ are conjugate and we use Gibbs sampling. The correlation parameters ρ_η and ρ_ξ and the smoothness parameters ν_η and ν_ξ are updated with Metropolis sampling, tuned to have acceptance ratio near 0·4. The sampling density parameters v_j are updated using blocked Metropolis sampling to account for posterior correlation between coefficients for nearby knots. We used ten blocks, with knots allocated to blocks using k-means clustering implemented by the `kmeans` package in R. For the simulation study in §5 we generated 5,000 samples and discarded the first 1,000 as burn-in. For the analysis of the ozone data in §6 we generated 20,000 samples and discarded the first 5,000. Convergence was monitored using trace plots of the deviance as well as several representative parameters.

5. SIMULATION STUDY

We conduct a simulation study to illustrate the effect of failing to account for informative sampling on spatial interpolation, and determine the amount of data need to reliably identify informative sampling. We assume $\mathcal{D} = [0, 1]^2$ and no spatial covariates, $x(s) = 1$ for all s. We generate data using model (7) with an equally-spaced grid of $N = 225$ knots on [-0·2,1·2]^2 and a Matérn kernel. We generate $S = 50$ data sets from each of four simulation scenarios: (1) $n = 250$, $a = 0$, ρ =0·2; (2) $n = 250$, $a = 1$, ρ =0·2; (3) $n = 250$, $a = 1$, ρ =0·5; and (4) $n = 500$, $a = 1$, ρ =0·2, with $\sigma = 1$, $E\{\mu(s)\} = 0$, ν =2·0, and τ =0·1 under all scenarios. For each simulated data set we fit the following three models. The noninformative sampling model sets $a = 0$, the plug-in model sets $\xi(s) = \hat{\xi}(s)$ to account for informative locations, and the full model implements the approach of §4. In the plug-in analysis the location density is estimated using kernel density estimation in R's `KernSur` function in the `GenKern` package with default settings. `GenKern` gives a bivariate kernel density estimate that uses Gaussian kernels with bandwidth chosen using a direct plug-in approach to approximate the asymptotically optimal bandwidth.

We use the same grid of $N = 225$ knots used to generate the data in the kernel convolution model, and approximate the integral using a square grid of $M = 900$ points $t_1, \ldots, t_M$ covering $[0, 1]$. Motivated by Rodrigues & Diggle (2010), we used an equally spaced grid of 225 knots on $[-0.2, 1.2]^2$. Simulation study results show that irrespective of the number and position of the sampling locations, the Gaussian process can be well approximated with 225 knots. Following Lee et al. (2005), the grid spacings are chosen to be no larger than the standard deviation of the kernel in the convolution representation. We use diffuse normal priors for β^* and a and the covariance parameters have priors $\sigma^2, \tau_\xi^2, \tau_\eta^2 \sim$ Inv-Ga$(0{\cdot}01, 0{\cdot}01)$, $\rho_\xi^2, \rho_\eta^2 \sim$ U$(0, 2)$, and $\nu_\xi^2, \nu_\eta^2 \sim$ U$(0, 30)$.

Table 1 reports bias, mean squared error, mean absolute deviation and coverage probability, each averaged over the grid of M spatial locations $t_1, \ldots, t_M$. The coverage probability is the proportion of the M grid locations for which the posterior 95% interval for $\mu(t_j)$ covers the true value. For the plug-in model and the full model we also report the power for a in Table 1 which is defined to be the proportion of data sets for which the posterior 95% credible interval for a excludes zero.

All three methods perform similarly when sampling is not informative. In this case, the informative sampling methods rarely identify a as significant and reduce to the usual geostatistical model. The noninformative sampling model has high mean squared error and negative bias in the remaining designs with informative sampling. The two methods that allow for informative sampling reduce mean squared error compared to the noninformative sampling model. The informative sampling models also reduce bias, although some bias remains, especially for design 2. In all cases the full model improves on the plug-in approach. The relative mean squared error of the noninformative sampling model to the full model is smaller for design 3 (0·132/0·108 = 1·222) with large spatial range and design 4 (0·256/0·190=1·347) with large sample size than for design 2 (0·494/0·329 = 1·502), so it seems that accounting for informative sampling is most important for small data sets with considerable spatial variation.

Table 1. *Simulation study results*

Design	Model	MSE($\times 10^2$)	MAD($\times 10^2$)	Bias($\times 10^2$)	CP($\times 10^2$)	Power for a($\times 10^2$)
1	NIS	33·1 (2·8)	41·3 (0·6)	2·0 (1·3)	93·0 (1·0)	–
	Plug-in	32·2 (1·7)	41·3 (0·)	2·5 (1·3)	93·0 (1·0)	10·0
	Full	31·9 (1·2)	41·5 (0·7)	2·5 (1·3)	93·0 (1·0)	10·0
2	NIS	49·4 (5·0)	50·0 (1·1)	−25·8 (1·3)	90·0 (1·0)	–
	Plug-in	39·2 (5·5)	44·8 (0·9)	−13·9 (1·3)	91·0 (1·0)	74·0
	Full	32·9 (2·8)	43·2 (0·8)	−7·5 (1·6)	93·0 (1·0)	80·0
3	NIS	13·2 (1·1)	28·1 (1·8)	−8·3 (1·4)	94·0 (1·0)	–
	Plug-in	12·1 (0·8)	27·1 (1·8)	−3·1 (1·4)	94·0 (1·0)	40·0
	Full	10·8 (0·7)	25·3 (1·4)	−2·0 (1·3)	95·0 (1·0)	50·0
4	NIS	25·6 (1·1)	36·9 (0·7)	−15·3 (1·2)	92·0 (1·0)	–
	Plug-in	20·9 (0·8)	33·9 (0·5)	−7·2 (1·1)	92·0 (1·0)	88·0
	Full	19·1 (0·6)	32·6 (0·4)	−0·8 (1·0)	94·0 (1·0)	98·0

NIS, noninformative sampling; MSE, mean squared error; MAD, mean absolute deviation; CP, coverage probability

To analyze sensitivity to the prior for a, we redid simulation design 2 with $a = 1$ and ρ=0·2 and used four different priors for a: $\mathrm{N}(1, 1)$, $\mathrm{N}(0, 1)$, $\mathrm{N}(0, 10^2)$ and an improper prior. In summary, mean squared prediction error and predictive coverage are insensitive to the hyperparameters of the prior on a for $n = 150$ and $n = 200$. Even for a sample size as small as $n = 50$, differences are small for different priors. However, the $\mathrm{N}(0, 10^2)$ prior and the informative prior $\mathrm{N}(1, 1)$ lead to a better power for a than the others when $n = 50$ and 100. The minimum sample size needed to swamp out the prior for a is around 150 in this example.

6. Analysis of Eastern United States ozone data

With the increasing concern about air pollution and climate change, building predictive models for ozone is an important area. It is often the case that the monitoring locations are informative about the ozone surface and hence it is important to account for informative sampling. We analyze the median daily ozone for June-August 2007 for $n = 631$ observations in the Eastern United States. The data are plotted in Fig. 1a. There is a clear association between the sampling density and the response, as there are more monitors placed in areas with high ozone, such as Atlanta and New England, than areas with low ozone, such as Mississippi and West Virginia. We fit a generalized additive model to the median ozone values and the kernel density estimate of the log sampling density using locally weighted scatterplot smoothing in Fig. 1b. The linear fit is entirely contained within the generalized additive model 95% confidence intervals for all values of the log sampling density estimate, supporting the log linear model in (1).

To apply a stationary spatial model we first project the spatial locations to a two-dimensional surface using the Mercator projection, and then scale them to the unit square coordinate-wise by subtracting the minimum and dividing by the range of the observation locations. We fit the informative sampling model with a 30 × 30 grid of knots on $[-0{\cdot}2, 1{\cdot}2]^2$ in the kernel convolution approximation in (6) and a 50 × 50 grid of points on $[0, 1]^2$ in the integral approximation in the sampling density (4). Points outside the convex hull of the observation locations or outside the continental United States were discarded from integral approximation to the sampling density, leaving $M = 1077$. Kernel convolution knots not within 0·1 of an integral approximation knot were discarded, leaving $N = 490$.

Table 2. *Mean and 95% intervals for the ozone data*

Parameters	NIS	Plug-in	Full
a	–	4·43 (2·16, 6·46)	3·21 (2·12, 4·25)
σ	4·68 (4·37, 5·03)	4·70 (4·38, 5·04)	4.78 (4.47, 5.12)
τ_g	0·17 (0·14, 0·27)	0·15 (0·13, 0·19)	0·17 (0·13, 0·21)
ρ_g	0·06 (0·05, 0·16)	0·06 (0·04, 0·10)	0·06 (0·05, 0·10)
ν_g	3·95 (0·92, 6·42)	3·46 (1·53, 5·52)	12·6 (0·74, 28·8)
τ_f	–	–	0·05 (0·04, 0·06)
ρ_f	–	–	0·07 (0·04, 0·13)
ν_f	–	–	10·7 (0·74, 28·77)

NIS, noninformative sampling

We include a second-order spatial trend as predictors in $x(s)$, that is, linear and quadratic terms for re-scaled latitude and longitude, and their interaction. We compare the noninformative sampling, plug-in and full models described in §5. The posteriors for several parameters are summarized in Table 2. The spatial process for both the mean process and sampling density are fairly smooth. The posterior 95% intervals for ν_ξ and ν_η exclude the exponential covariance ($\nu = 0{\cdot}5$) for all three models.

The 95% interval of a for both the plug-in model (2·16, 6·46) and fully Bayesian model (2·12, 4·25) excludes zero, indicating an informative sampling scheme. The scale of a's posterior is not comparable between the two models since the plug-in density estimate has been standardized to have mean zero and variance one. The effect of accounting for informative sampling is illustrated in Fig. 2. The difference in predicted values between the noninformative sampling and full model in Fig. 2c is the largest in Northern Pennsylvania and West Virginia. These areas have relatively few monitors and are near areas with high ozone. The difference between the noninformative sampling and plug-in predictions in Fig. 2d are also positive in these areas though the differences are not nearly as large in the plug-in analysis. This may be because the plug-in estimates do not appropriately account for uncertainty in estimation, and hence may lead to some attenuation of the estimated surface.

Finally, we refit the model with different priors and different knot locations to test for sensitivity to these assumptions. We fit the model with 20×20 and 40×40 initial grids of knots in the kernel convolution approximation. After removing knots outside the domain of interest, this gave $N = 206$ and $N = 876$ knots, respectively. The results were fairly similar to the original 30×30 grid. In all cases the posterior of a was separated from zero, the posterior median being 3·31 and 2·85 for $N = 206$ and $N = 876$ knots, respectively, and the largest difference between the noninformative sampling and full model was in the Northern Pennsylvania and West Virginia.

7. Discussion

We have focused on a simple model for informative locations, which assumes that the outcomes are conditionally independent of the locations given the mean process $\mu(s)$ and the spatial location density $p(s)$. In addition, we include a single parameter a controlling the informativeness of the sampling process. These simplifying assumptions certainly make the theory and computation more tractable. However, to more realistically characterize data from a broader variety of applications, it may be necessary to generalize the models. There are several interesting directions in this regard. First, it is straightforward conceptually to replace the constant a with a spatially-varying coefficient $a(s)$, which is assigned a Gaussian process prior. This generalization allows the informativeness of the sampling locations to vary spatially; for example, in certain regions, say near cities, monitors may be placed without regard to the outcome, while in other regions, say in the rural areas, monitors may be placed at sites likely to have high values. It is an open question whether one can consistently estimate $a(s)$ in this extended model without very restrictive assumptions. However, a simple adjustment for informative sampling may be preferable to more complicated models that require rich datasets for reliable estimation.

ACKNOWLEDGEMENT

This research was partially supported by the National Institute of Environmental Health Sciences of the National Institutes of Health. The authors would like to thank Mr. Avishek Chakraborty and Mr. Anirban Bhattacharya for their helpful comments.

APPENDIX

Proof of Theorem 1

Let $\phi = (\xi_r, \eta_r, \beta_\xi, \beta_\eta, a, \sigma)$ and $\phi_0 = (\xi_{0r}, \eta_{0r}, \beta_{\xi 0}, \beta_{\eta 0}, a_0, \sigma_0)$ be a fixed set of parameters in $\mathcal{C}(\mathcal{D}) \times \mathcal{C}(\mathcal{D}) \times \mathbb{R} \times \mathbb{R}^+$. Clearly $(y_i, s_i) \sim f(y, s \mid \phi)$, where

$$f(y, s \mid \phi) = f(y \mid s, \phi) p(s \mid \phi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\{y - \mu(s)\}^2}{2\sigma^2}\right] \frac{\exp\{x(s)^{\mathrm{T}}\beta_\xi + \xi_r(s)\}}{\int_{\mathcal{D}} \exp\{x(s)^{\mathrm{T}}\beta_\xi + \xi_r(s)\} ds}.$$

Here $\mu(s) = x(s)^{\mathrm{T}}(a\beta_\xi + \beta_\eta) + a\xi_r(s) + \eta_r(s)$. Let $\mu_0(s) = x(s)^{\mathrm{T}}(a_0\beta_{\xi 0} + \beta_{\eta 0}) + a_0\xi_{0r}(s) + \eta_{0r}(s)$. Define $\Lambda(\phi_0, \phi) = \log\big\{f(y, s \mid \phi_0)/f(y, s \mid \phi)\big\}$ and $K(\phi_0, \phi) = E_{\phi_0}\{\Lambda(\phi_0, \phi)\}$. Then following Schwartz (1965), its enough to show that for all $\epsilon > 0$,

$$(\Pi_{\xi_r} \times \Pi_{\eta_r} \times \pi_{\beta_\xi} \times \pi_{\beta_\eta} \times \pi_\sigma \times \pi_a)\big\{\phi : K(\phi_0, \phi) < \epsilon\big\} > 0.$$

We calculate $K(\phi_0, \phi)$ below.

$$\begin{aligned}
K(\phi_0, \phi) &= E_{\phi_0}\{\Lambda(\phi_0, \phi)\} = E_{\phi_0}\left\{\log \frac{f(y, s \mid \phi_0)}{f(y, s \mid \phi)}\right\} \\
&= \frac{1}{2}\log\frac{\sigma^2}{\sigma_0^2} + E_{\phi_0}\left[-\frac{\{y - \mu_0(s)\}^2}{2\sigma_0^2}\right] - E_{\phi_0}\left[-\frac{\{y - \mu(s)\}^2}{2\sigma^2}\right] - \\
&\qquad E_{\phi_0}\big\{x(s)^{\mathrm{T}}(\beta_\xi - \beta_{\xi 0}) + \xi_r(s) - \xi_{0r}(s)\big\} + \log\left[\frac{\int_{\mathcal{D}} \exp\big\{x(s)^{\mathrm{T}}\beta_\xi + \xi_r(s)\big\} ds}{\int_{\mathcal{D}} \exp\big\{x(s)^{\mathrm{T}}\beta_{\xi 0} + \xi_{0r}(s)\big\} ds}\right] \\
&= \frac{1}{2}\log\frac{\sigma^2}{\sigma_0^2} - \frac{1}{2}\left(1 - \frac{\sigma_0^2}{\sigma^2}\right) + \frac{1}{2\sigma^2}\int_{\mathcal{D}} \{\mu_0(s) - \mu(s)\}^2 p(s) ds \\
&\quad + \int_{\mathcal{D}} \{x(s)^{\mathrm{T}}(\beta_\xi - \beta_{\xi 0}) + \xi_r(s) - \xi_{0r}(s)\} p(s) ds + \log\left[\frac{\int_{\mathcal{D}} \exp\big\{x(s)^{\mathrm{T}}\beta_\xi + \xi_r(s)\big\} ds}{\int_{\mathcal{D}} \exp\big\{x(s)^{\mathrm{T}}\beta_{\xi 0} + \xi_{0r}(s)\big\} ds}\right].
\end{aligned}$$

For each $\delta > 0$, define

$$B_\delta = \big\{\phi : ||\xi_r - \xi_{0r}||_\infty < \delta, ||\eta_r - \eta_{0r}||_\infty < \delta, ||\beta_\xi - \beta_{f0}|| < \delta, ||\beta_g - \beta_{g0}|| < \delta, |a - a_0| < \delta, \\ |\sigma/\sigma_0 - 1| < \delta\big\}.$$

Take $b_1 = ||\mu_0 - \mu||_\infty$ and $b_2 = \sigma/\sigma_0$. Let $g_1(b_1, b_2) = \log b_2 - (b_2^2 - 1)/(2b_2^2) + b_1^2/(2\sigma_0^2 b_2^2)$. Clearly $g_1(b_1, b_2)$ is continuous at $b_1 = 0$ and $b_2 = 1$ and $g_1(0, 1) = 0$. We have

$$b_1 \leq M||(a\beta_\xi + \beta_\eta) - (a_0\beta_{\xi 0} + \beta_{\eta 0})|| + ||\{a\xi_r(s) + \eta_r(s)\} - \{a_0\xi_{0r}(s) + \eta_{0r}(s)\}||$$

and

$$K(\phi_0,\phi) \le g_1(b_1,b_2) + \int_{\mathcal{D}} \big\{x(s)^{\mathrm{T}}(\beta_\xi - \beta_{\xi 0}) + \xi_r(s) - \xi_{0r}(s)\big\} p(s)ds$$
$$+ \log\left[\frac{\int_{\mathcal{D}} \exp\big\{x(s)^{\mathrm{T}}\beta_\xi + \xi_r(s)\big\}ds}{\int_{\mathcal{D}} \exp\big\{x(s)^{\mathrm{T}}\beta_{\xi 0} + \xi_{0r}(s)\big\}ds}\right].$$

For $\epsilon > 0$, there exists a $\delta_1 > 0$ such that for all $\phi \in B_{\delta_1}$,

$$\frac{1}{2}\log\frac{\sigma^2}{\sigma_0^2} - \frac{1}{2}\left(1 - \frac{\sigma_0^2}{\sigma^2}\right) + \frac{1}{2\sigma^2}\int_{\mathcal{D}}\{\mu_0(s) - \mu(s)\}^2 p(s)ds < \frac{\epsilon}{3}.$$

Also there exists $\delta_2 > 0$ such that for all $\phi \in B_{\delta_2}$, $\big\{x(s)^{\mathrm{T}}(\beta_\xi - \beta_{\xi 0}) + \xi_r(s) - \xi_{0r}(s)\big\} < \epsilon/3$ uniformly for all $s \in \mathcal{D}$. If we define $h_\phi(s) = \exp\big\{x(s)^{\mathrm{T}}\beta_\xi + \xi_r(s)\big\}$, then $\phi \mapsto \int_{\mathcal{D}} h_\phi(s)ds$ is a continuous function and hence $\phi \mapsto \log\big\{\int_{\mathcal{D}} h_\phi(s)ds\big\}$ is also a continuous function. So there exists a $\delta_3 > 0$ such that

$$\phi \in B_{\delta_3} \Rightarrow \log\left\{\int_{\mathcal{D}} h_\phi(s)ds\right\} - \log\left\{\int_{\mathcal{D}} h_{\phi_0}(s)ds\right\} < \frac{\epsilon}{3}.$$

Choosing $\delta = \min\{\delta_1, \delta_2, \delta_3\}$, $\phi \in B_\delta$ implies $K(\phi_0, \phi) < \epsilon$. From Choi (2005), it follows that with the priors specified in §3·1

$$(\Pi_{\xi_r} \times \Pi_{\eta_r} \times \pi_{\beta_\xi} \times \pi_{\beta_\eta} \times \pi_\sigma \times \pi_a)(B_\delta) > 0.$$

Hence,

$$(\Pi_{\xi_r} \times \Pi_{\eta_r} \times \pi_{\beta_\xi} \times \pi_{\beta_\eta} \times \pi_\sigma \times \pi_a)\big\{\phi : K(\phi_0,\phi) < \epsilon\big\} > 0.$$

Proof of Theorem 2

The prior specifications on $\rho_\xi, \rho_\eta, \tau_\xi$ and τ_η enable one to bound any quadratic forms and determinants involving Σ_ξ^n and Σ_η^n by fixed quantities. Hence, in showing that the posterior $p(a \mid y, s)$ is proper, its enough to treat $\rho_\xi, \rho_\eta, \tau_\xi$ and τ_η as constants. Without loss of generality we can work with $\mathcal{D} = [0,1]^2$ by the projection argument described in §6. Following Benes et al. (2003), we consider the grid approximation of the infinite dimensional Gaussian process $\{\xi_r(s) : s \in \mathcal{D}\}$, denoted by ξ_r. Let $\mathcal{D} = \bigcup_{j=1}^J I_j$, with $\{I_j\}$ denoting a segmentation of $\mathcal{D}$ into contiguous regions of equal area $\Delta = J^{-1}\int_{\mathcal{D}} ds$. Choose J sufficiently large such that at most one s_i lies within any I_j. The infinite-dimensional Gaussian process, ξ_r, can be approximated by a finite dimensional vector $\xi_r^J = (\xi_r^{*1}, \ldots, \xi_r^{*J})^{\mathrm{T}}$, corresponding to the choice of arbitrary points $s_1^*, \ldots, s_J^*$ within $I_1, \ldots, I_J$, respectively such that $\xi_r(s_i) = \xi_r^{*j}$ if $s_i \in I_j$. Thus $\xi_r^J \sim \mathrm{N}(0, \Sigma_\xi^{*J})$, where $(\Sigma_\xi^{*J})_{ij} = c(||s_i^* - s_j^*|| \mid \psi)$. Define the true posterior $p^{true}(\xi_r^n \mid s)$ and the approximated posterior $p^J(\xi_r^n \mid s)$ as follows.

$$p^{true}(\xi_r^n \mid s) \propto p^{true}(\xi_r^n, s) = E\left\{\int \exp\big\{x(s)^{\mathrm{T}}\beta_\xi + \xi_r(s)\big\}ds \mid \xi_r^n\right\}^{-n} \exp\big\{-0{\cdot}5\big(\xi_r^n\big)^{\mathrm{T}}\big(\Sigma_\xi^n\big)^{-1}\big(\xi_r^n\big)^{\mathrm{T}}\big\}$$

and

$$p^J(\xi_r^n \mid s) \propto p^J(\xi_r^n, s) = \left[\Delta\sum_{j=1}^J \exp\big\{x(s_j^*)^{\mathrm{T}}\beta_\xi + \xi_r(s_j^*)\big\}\right]^{-n} \exp\big\{-0{\cdot}5\big(\xi_r^J\big)^{\mathrm{T}}\big(\Sigma_\xi^{*J}\big)^{-1}\big(\xi_r^J\big)^{\mathrm{T}}\big\}.$$

Marginalizing out η_r^n, we have $y \mid s, \xi_r, a, \sigma^2, \beta_\eta, \beta_\xi \sim \mathrm{N}(X\beta^* + a\xi_r^n, \sigma^2 I_n + \Sigma_\eta^n)$, where $X^{\mathrm{T}} = \{x(s_1)\cdots x(s_n)\}$. The true posterior of $(\xi_r^n, a, \sigma^2, \beta_\xi, \beta_\eta)$ is

$$p^{true}(\xi_r^n, a, \beta_\xi, \beta_\eta, \sigma^2 \mid y, s) \propto p(y \mid s, \xi_r, a, \sigma^2, \beta_\xi, \beta_\eta)p^{true}(\xi_r^n, s)\pi(\sigma^2)\pi(\beta_\xi)\pi(\beta_\eta).$$

Benes et al. (2003) showed that, under these assumptions, for a fixed $s \in \mathcal{D}^n$, the expectation of any bounded function with respect to $p^J(\xi_r^n \mid s)$ converges to the corresponding expectation with respect to $p^{true}(\xi_r^n \mid s)$ as J tends to infinity. Hence there exists a J such that the expectation of the bounded function with respect to $p^J(\xi_r^n \mid s)$ is greater than the corresponding expectation with respect to $(1/2)p^{true}(\xi_r^n \mid s)$. Thus, in order to show propriety of the true posterior of $(\xi_r^n, a, \sigma^2, \beta_\xi, \beta_g)$, which involves $p^{true}(\xi_r^n \mid s)$, its enough to show the propriety of the approximated posterior $p^J(\xi_r^n, a, \beta_\xi, \beta_\eta, \sigma^2 \mid y, s)$. The approximated posterior of $(\xi_r^n, a, \sigma^2, \beta_\xi, \beta_\eta)$ is

$$p^J(\xi_r^n, a, \beta_\xi, \beta_\eta, \sigma^2 \mid y, s) = C \exp\big\{-0{\cdot}5\big(Y - X\beta^* - a\xi_r^n\big)^{\mathrm{T}}\big(\sigma^2 I_n + \Sigma_\eta^n\big)^{-1}\big(Y - X\beta^* - a\xi_r^n\big)\big\}\times$$

$$\exp\big\{-0{\cdot}5\big(\xi_r^J\big)^{\mathrm{T}}\big(\Sigma_\xi^{*J}\big)^{-1}\big(\xi_r^J\big)\big\}\pi(\beta_\xi)\pi(\beta_\eta)\times\pi(\sigma^2)\frac{\exp\big\{\sum_{i=1}^n x(s_i)^{\mathrm{T}}\beta_\xi + \xi_r(s_i)\big\}}{\Delta^n\big[\sum_{j=1}^J \exp\big\{x(s_j^*)^{\mathrm{T}}\beta_\xi + \xi_r^{*j}\big\}\big]^n},$$

where C is a constant. Since $\exp\big\{x(s_i)^{\mathrm{T}}\beta_\xi + \xi_r(s_i)\big\} < \sum_{j=1}^J \exp\big\{x(s_j^*)^{\mathrm{T}}\beta_\xi + \xi_r^{*j}\big\}$ for all $i = 1, \ldots, n$,

$$\frac{\exp\big\{\sum_{i=1}^n x(s_i)^{\mathrm{T}}\beta_\xi + \xi_r(s_i)\big\}}{\big[\sum_{j=1}^J \exp\big\{x(s_j^*)^{\mathrm{T}}\beta_\xi + \xi_r^{*j}\big\}\big]^n} < 1.$$

After integrating out ξ_r^J excluding ξ_r^n we are left with

$$p(\xi_r^n, a, \beta_\xi, \beta_\eta, \sigma^2 \mid Y, s) \le C_1 \exp\big\{-0{\cdot}5\big(Y - X\beta^* - a\xi_r^n\big)^{\mathrm{T}}\big(\sigma^2 I_n + \Sigma_\eta^n\big)^{-1}\big(Y - X\beta^* - a\xi_r^n\big)\big\}\times$$

$$\exp\big\{-0{\cdot}5\big(\xi_r^n\big)^{\mathrm{T}}\big(\Sigma_f^n\big)^{-1}\big(\xi_r^n\big)\big\}\pi(\beta_\xi)\pi(\beta_\eta)\pi(\sigma^2),$$

where $C_1 > 0$ is a constant and Σ_ξ^n is the variance-covariance matrix of ξ_r^n constructed out of Σ_ξ^{*J}. Setting $Z = \big(y - X\beta^*\big)/a$, $\Sigma = \big(\Sigma_\eta^n + \sigma^2 I_n\big)/a^2$ and $\Omega_\eta = \big\{(\Sigma_\xi^n)^{-1} + \Sigma^{-1}\big\}^{-1}$ and completing quadratic forms yield

$$p(\xi_r^n, a, \beta_\xi, \beta_\eta, \sigma^2 \mid y, s) \le C_2 \exp\big\{-0{\cdot}5\big(\xi_r^n - \Omega_\eta\Sigma^{-1}Z\big)^{\mathrm{T}}\Omega_\eta^{-1}\big(\xi_r^n - \Omega_\eta\Sigma^{-1}Z\big)\big\}\times$$

$$\exp\big\{-0{\cdot}5\big(Z^{\mathrm{T}}\Sigma^{-1}Z - Z^{\mathrm{T}}\Sigma^{-1}\Omega_\eta\Sigma^{-1}Z\big)\big\}\pi(\beta_\xi)\pi(\beta_\eta)\pi(\sigma^2),$$

where $C_2 > 0$ is another constant. Next we state a useful lemma from matrix algebra.

LEMMA 1. *If A and B are positive definite square matrices so is $A - A(A+B)^{-1}A$.*

Proof. We have

$$A - A(A+B)^{-1}A = A(A+B)^{-1}B = \{B^{-1}(A+B)A^{-1}\}^{-1} = (B^{-1} + A^{-1})^{-1}.$$

The conclusion follows from the fact that the sum and inverses of positive definite matrices of the same dimension are also positive definite. □

From Lemma 1, we have $\big(Z^{\mathrm{T}}\Sigma^{-1}Z - Z^{\mathrm{T}}\Sigma^{-1}\Omega_\eta\Sigma^{-1}Z\big) \ge 0$, so that

$$p(\xi_r^n, a, \beta_\xi, \beta_\eta, \sigma^2 \mid y, s) \le C_2 \exp\big\{-0{\cdot}5\big(\xi_r^n - \Omega_\eta\Sigma^{-1}Z\big)^{\mathrm{T}}\Omega_\eta^{-1}\big(\xi_r^n - \Omega_\eta\Sigma^{-1}Z\big)\big\}\times$$

$$\pi(\beta_\xi)\pi(\beta_\eta)\pi(\sigma^2).$$

Integrating out ξ_r^n first and then β_ξ and β_η,

$$p(a, \sigma^2 \mid y, s) \le C_3\big|\big(\Sigma_\xi^n\big)^{-1} + a^2\big(\Sigma_\eta^n + \sigma^2 I_n\big)^{-1}\big|^{-(1/2)}.$$

Call $\Sigma_\xi^n = A$ and $\Sigma_\eta^n = B$. Hence

$$\big|A^{-1} + a^2(B + \sigma^2 I)^{-1}\big| = \frac{|I + a^2 A(B + \sigma^2 I)^{-1}|}{|A|} = \frac{|a^2 A + \sigma^2 I + B|}{|\sigma^2 I + B||A|}.$$

Now we state a useful result from matrix algebra.

PROPOSITION 1. *If A and B and non-negative definite matrices, then $|A+B| \geq |A| + |B|$ with strict inequality holding in case of positive definite matrices.*

Using Proposition 1, we get

$$\left(\frac{|a^2A+\sigma^2I+B|}{|\sigma^2I+B|}\right)^{-(1/2)} \leq \left(\frac{|a^2A|+|\sigma^2I+B|}{|\sigma^2I+B|}\right)^{-(1/2)} = \left\{1+\frac{a^{2n}|A|}{\prod_{i=1}^{n}(\sigma^2+b_i)}\right\}^{-(1/2)} \leq \left\{1+\frac{a^{2n}|A|}{(\sigma^2+b_n)^n}\right\}^{-(1/2)},$$

where $0 < b_1 \leq b_2 \leq \cdots \leq b_n$ are the eigen values of B. By Minkowski's inequality we get

$$\left\{1+\frac{a^{2n}|A|}{(\sigma^2+b_n)^n}\right\}^{-(1/2)} \leq \frac{(\sigma^2+b_n)^{n/2}}{c_n\left(a^2|A|^{(1/n)}+\sigma^2+b_n\right)^{n/2}}.$$

Set $|A|^{1/n} = k_1$ and $b_n = k_2$. We assume $n \geq 2$. Then ignoring constants

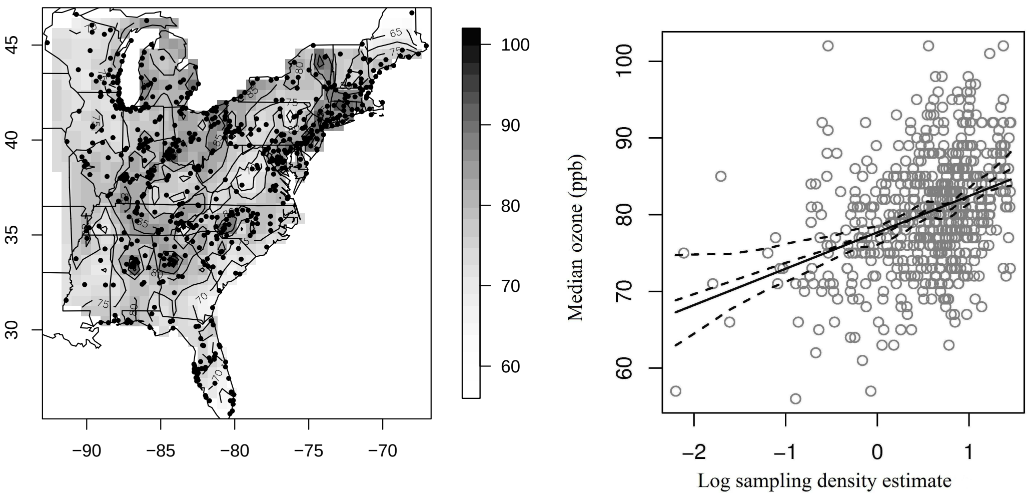
$$\int_{-\infty}^{\infty} \frac{(\sigma^2+b_n)^{n/2}}{\left(a^2|A|^{1/n}+\sigma^2+b_n\right)^{n/2}}da = \int_{-\infty}^{\infty} \frac{1}{\left\{1+(a^2k_1)/(\sigma^2+k_2)\right\}^{n/2}}da \leq \int_{-\infty}^{\infty} \frac{1}{\left\{1+(a^2k_1)/(\sigma^2+k_2)\right\}}da = \pi\left(\frac{\sigma^2+k_2}{k_1}\right)^{(1/2)}.$$

Now since $E_\pi(\sigma) < \infty$,

$$\int_0^{\infty}\left(\frac{\sigma^2+k_2}{k_1}\right)^{(1/2)}\pi(d\sigma^2) < \infty.$$

By Fubini's Theorem, $p(a \mid Y, s)$ is integrable.

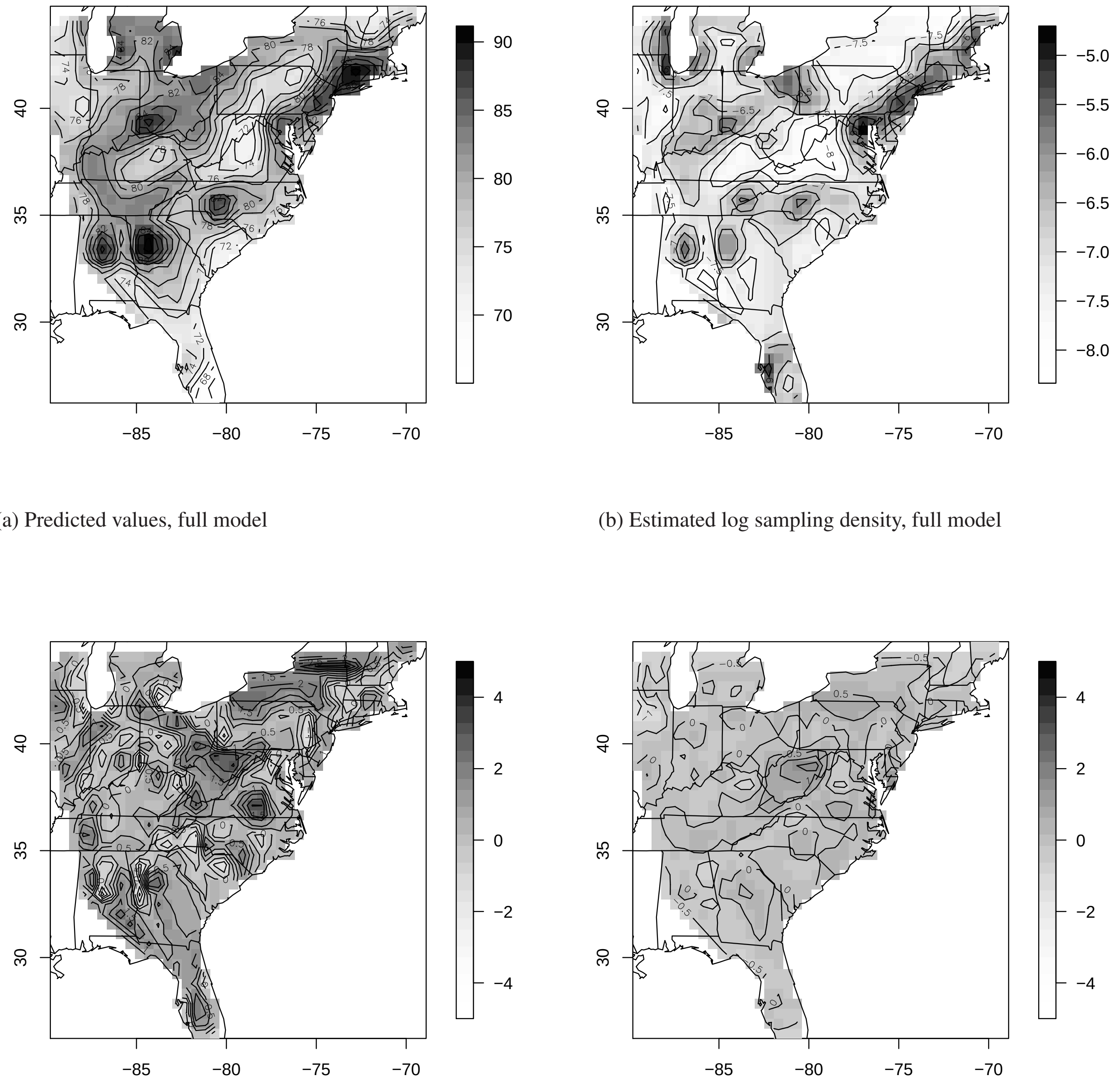
Fig. 1. Plots of the ozone data. Panel (a) plots the ozone data (ppb; color) and monitor locations (points), Panel (b) plots the estimated log sampling density against the response.



(a) Median ozone

(b) Log sampling density versus median ozone (circles), gamfit with 95% intervals (dashed), linear fit (solid)

Fig. 2. Posterior mean predicted values (Panel (a)) and log sampling density (Panel (b)) from the full model, and the difference in posterior mean predicted values from the non-informative sampling model and the full model (Panel (c)) and plug-in model (Panel (d))



(a) Predicted values, full model

(b) Estimated log sampling density, full model

(c) Predicted values, Noninformative sampling − full model

(d) Predicted values, Noninformative sampling − plug-in model

REFERENCES

BENES, V., BODLÁK, K., MØLLER, J. & WAAGEPETERSEN, R. P. (2003). Bayesian analysis of log Gaussian Cox process models for disease mapping. In *The ISI International Conference on Environmental Statistics and Health, Univ Santiago de Compostela*.

CHOI, T. (2005). *Posterior Consistency in Nonparametric Regression problems in Gaussian Process Priors*. Ph.D. thesis, Department of Statistics, Carnegie Mellon University.

CHOI, T. (2007). Alternative posterior consistency results in nonparametric binary regression using Gaussian process priors. *Journal of Statistical Planning and Inference* **137**, 2975–2983.

CHOI, T. & SCHERVISH, M. (2007). On posterior consistency in nonparametric regression problems. *Journal of Multivariate Analysis* **98**, 1969–1987.

DIGGLE, P., MENEZES, R. & SU, T. (2010). Geostatistical inference under preferential sampling (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **59**, 191–232.

HIGDON, D. (2002). Space and space-time modeling using process convolutions. *Quantitative methods for current environmental issues* , 37–56.

HO, L. & STOYAN, D. (2008). Modeling marked point patterns by intensity-marked Cox processes. *Statistics and Probability Letters* **78**, 1194–1199.

LEE, H., HIGDON, D., CALDER, C. & HOLLOMAN, C. (2005). Efficient models for correlated data via convolutions of intrinsic processes. *Statistical Modelling* **5**, 53–74.

MENEZES, R. (2005). *Assessing spatial dependency under non-standard sampling*. Ph.D. thesis, Universidad de Santiago de Compostela, Santiago de Compostela, Spain.

MØLLER, J., SYVERSVEEN, A. & WAAGEPETERSEN, R. (2001). Log Gaussian Cox processes. *Scandinavian Journal of Statistics* **25**, 451–482.

RADCLIFFE, S. J., GUO, W. & TEN HAVE, T. (2004). Joint modeling of longitudinal and survival data via a common frailty. *Biometrics* **60**, 892–899.

RODRIGUES, A. & DIGGLE, P. (2010). A Class of Convolution-Based Models for Spatio-Temporal Processes with Non-Separable Covariance Structure. *Scandinavian Journal of Statistics* (to appear).

SCHWARTZ, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete* **4**, 10–26.

STEIN, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer Series in Statistics.

VAN DER VAART, A. & VAN ZANTEN, J. (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse Gamma bandwidth. *The Annals of Statistics* **37**, 2655–2675.

WU, M. & FOLLMANN, D. (1999). Use of summary measures to adjust for informative missingness in repeated measures data with random effects. *Biometrics* **55**, 75–84.

ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association* **99**, 250–261.