Biostatistics (2001), **2**, 1, *pp.* 31–45
Printed in Great Britain

On the change of support problem for spatio-temporal data

ALAN E. GELFAND[†]
Department of Statistics, University of Connecticut, Storrs, Connecticut 06269, USA
Email: alan@stat.ucom.edu

LI ZHU, BRADLEY P. CARLIN
Division of Biostatistics, School of Public Health, University of Minnesota, Box 303 Mayo Building, Minneapolis, Minnesota 55455, USA

SUMMARY

In practice, spatial data are sometimes collected at points (i.e. point-referenced data) and at other times are associated with areal units (i.e. block data). The *change of support problem* is concerned with inference about the values of a variable at points or blocks different from those at which it has been observed. In the context of block data which can be sensibly viewed as averaging over point data, we propose a unifying approach for prediction from points to points, points to blocks, blocks to points, and blocks to blocks. The approach includes fully Bayesian kriging. We also extend our approach to the the case of spatio-temporal data, wherein a judicious specification of spatio-temporal association enables manageable computation. Exemplification of the static spatial case is provided using a dataset of point-level ozone measurements in the Atlanta, Georgia metropolitan area. The dynamic spatial case is illustrated using a temporally extended version of this dataset, enabling comparison at the common time point.

Keywords: Bayesian methods; Environmental risk analysis; Geographic Information System (GIS); Kriging; Modifiable areal unit problem; Simulation-based model fitting.

1. INTRODUCTION

Consider a univariate variable that is spatially observed. In particular, assume that it is observed either at points in space, which we refer to as *point-referenced* or simply *point* data, or over areal units (e.g. counties or zip codes), which we refer to as *block* data. The *change of support problem* is concerned with inference about the values of the variable at points or blocks different from those at which it has been observed.

In the case where the data are collected exclusively at blocks and inference is sought exclusively at new blocks, the problem has a rich literature and is often referred to as the *modifiable areal unit problem* (see, for example, Cressie (1996)). In the case of an *extensive* variable (i.e. one whose value for a block can be viewed as a sum of sub-block values, as in the case of population, disease counts, productivity or wealth), areal weighting offers a simple imputation strategy. While rather naive, such allocation proportional to area has a long history and is routinely available in geographic information system (GIS) software. Recognizing that areal allocation does not capture tendency toward spatial clustering, model-based approaches using 'better' covariates than area have recently emerge: see, for example, Flowerdew and Green (1992), Mugglin and Carlin (1998), Mugglin *et al.* (2000), and Zhu *et al.* (2000).

[†]To whom correspondence should be addressed.

1.1 *Motivating dataset*

A solution to the change of support problem is required in many health science applications, particularly spatial and environmental epidemiology. To illustrate, we consider a dataset of ozone levels in the Atlanta, GA metropolitan area, as reported by Tolbert *et al.* (2000). Ozone measures are available at between 8 and 10 fixed monitoring sites during the 92 summer days (June 1 to August 31) of 1995. Figure 1 shows the one-hour daily maximum ozone measures at the 10 monitoring sites on July 15, 1995, along with the boundaries of the 162 zip codes in the Atlanta metropolitan area. Here we might be interested in predicting the ozone level at different points on the map (say, the two points marked A and B, which lie on opposite sides of a single city zip), or the average ozone level over a particular zip (say, one of the 36 zips falling within the city of Atlanta, the collection of which are encircled by the dark boundary on the map). The latter problem is of special interest, since in this case relevant health outcome data are available only at the zip level. In particular, for each day and zip, we have the number of paediatric ER visits for asthma, as well as the total number of paediatric ER visits. Thus an investigation of the relationship between ozone exposure and paediatric asthma cannot be undertaken until the mismatch in the support of the two variables is resolved. Situations like this are relatively common, since personal privacy concerns often limit statisticians' access to health outcome data other than at the block average level.

A previous study of this dataset by Carlin *et al.* (1999) realigned the point-level ozone measures to the zip level by using an `ARC/INFO` universal kriging procedure to fit a smooth ozone exposure surface, and subsequently took the kriged value at each zip centroid as the ozone value for that zip. But this approach uses a single centroid value to represent the ozone level in the entire zip, and fails to properly capture variability and spatial association by treating these kriged estimates as observed values.

1.2 *Model assumptions and analytic goals*

The setting we work with is to assume that we are observing a continuous variable and that underlying all observations of this variable is a spatial process. We denote this process by $Y(\mathbf{s})$ for locations $\mathbf{s} \in D$, a region of interest. In our applications $D \subset \mathbb{R}^2$ but our development works in arbitrary dimensions. A realization of the process is a surface over D. For point-referenced data the realization is observed at a finite set of sites, say $\mathbf{s}_i, i = 1, 2, \ldots, I$. For block data we assume the observations arise as block averages. That is, for a block $B \subset D$,

$$Y(B) = |B|^{-1} \int_B Y(\mathbf{s})\, d\mathbf{s} \tag{1}$$

where $|B|$ denotes the area of B (see, for example, Cressie (1993)). The integration in (1) is an average of random variables, hence a random or stochastic integral. Thus, the assumption of an underlying spatial process is only appropriate for block data that can be sensibly viewed as an averaging over point data; examples of this would include rainfall, pollutant level, temperature and elevation. It would be inappropriate for, say, population, since there is no 'population' at a particular point. It would also be inappropriate for most proportions of interest. For instance, if $Y(B)$ is the proportion of college-educated persons in B, then $Y(B)$ is continuous but $Y(\mathbf{s})$, assuming a single individual at a point, is binary.

In general, we envision four possibilities. First, starting with point data $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_I)$, we seek to predict at new locations, i.e. to infer about $Y(\mathbf{s}'_1), \ldots, Y(\mathbf{s}'_K)$ (points to points). Second, starting with point data, we seek to predict at blocks, i.e. to infer about $Y(B_1), \ldots, Y(B_K)$ (points to blocks). Third, starting with block data $Y(B_1), \ldots, Y(B_I)$, we seek to predict at a set of locations, i.e. to infer about $Y(\mathbf{s}'_1), \ldots, Y(\mathbf{s}'_K)$ (blocks to points). Finally, starting with block data, we seek to predict at new blocks, i.e. to infer about $Y(B'_1), \ldots, Y(B'_K)$ (blocks to blocks).

All of this prediction may be collected under the umbrella of kriging (again, see Cressie (1993)). Our kriging will be implemented within the Bayesian framework enabling full inference (a posterior predictive

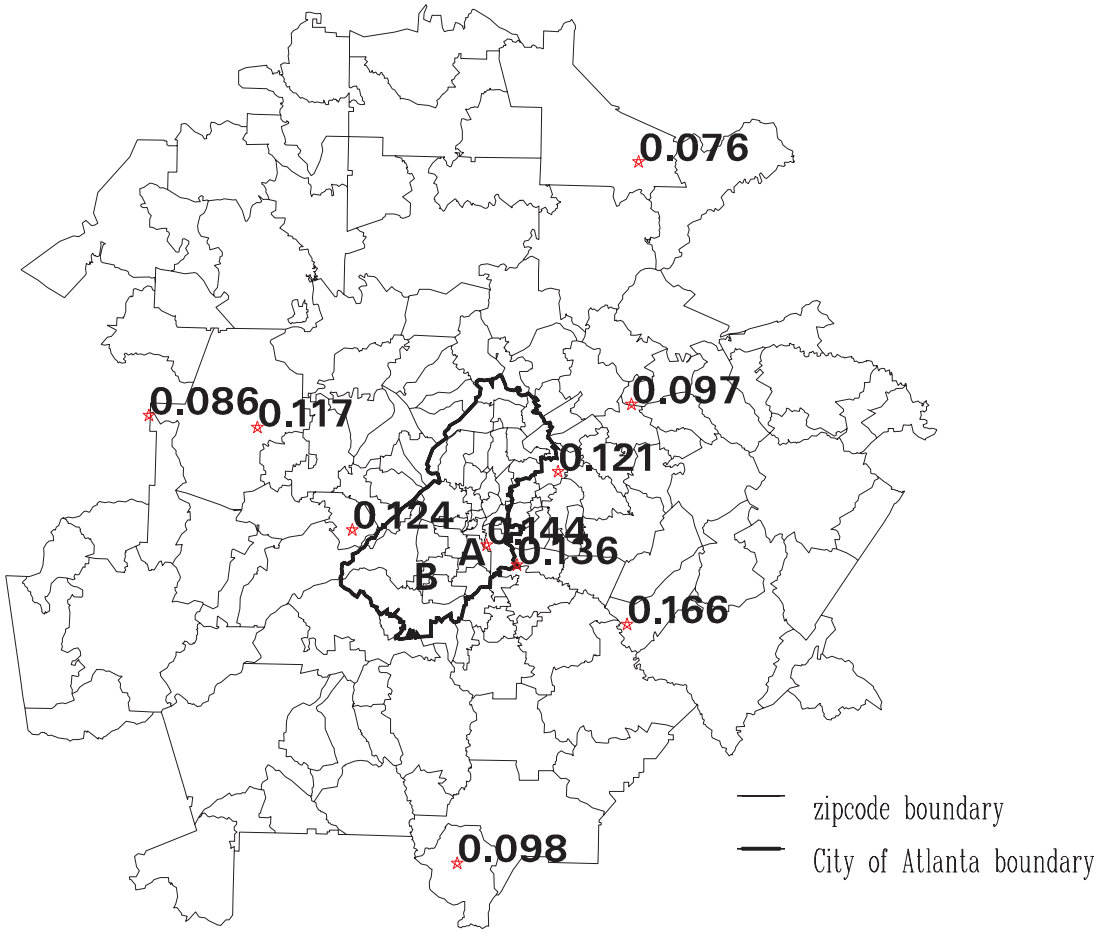

Fig. 1. Zip code boundaries in the Atlanta metropolitan area and ozone levels at the 10 monitoring sites for July 15, 1995.

distribution for every prediction of interest, joint distributions for all pairs of predictions, etc.) and avoiding asymptotics. We will, however, use rather noninformative priors, so that our results will roughly resemble those of a likelihood analysis.

There is a substantial body of literature focusing on spatial prediction from a Bayesian perspective. This includes Le and Zidek (1992), Handcock and Stein (1993), Brown *et al.* (1994), Handcock and Wallis (1994), DeOliveira *et al.* (1997), Ecker and Gelfand (1997), Diggle *et al.* (1998), and Karson *et al.* (1999). The work of Woodbury (1989), Abrahamsen (1993) and Omre and colleagues (Omre, 1987, 1988; Omre and Halvorsen, 1989; Omre *et al.*, 1989; Hjort and Omre, 1994) is partially Bayesian in the sense that prior specification of the mean parameters and covariance function are elicited; however, no distributional assumption is made for the $Y(\mathbf{s})$.

Inference about blocks through averages as in (1) is not only formally attractive but demonstrably preferable to *ad hoc* approaches. One such approach would be to average over the observed $Y(\mathbf{s}_i)$ in B. But this presumes there is at least one observation in any B, and ignores the information about the spatial process in the observations outside of B. Another *ad hoc* approach would be to simply predict the value

at some central point of B. But this value has larger variability than (and may be biased for) the block average.

A new wrinkle we add to the change of support problem here is application in the context of spatio-temporal data. In our Atlanta dataset, the number of monitoring stations is small but the amount of data collected over time is substantial. In this case, under suitable modelling assumptions, we may not only learn about the temporal nature of the data but also enhance our understanding of the spatial process. Moreover, the additional computational burden to analyse the much larger dataset within the Bayesian framework turns out still to be manageable.

More specifically, we illustrate with point-referenced data, assuming observations $Y(s_i, t_j)$ at locations s_i, $i = 1, \ldots, I$, and at times t_j, $j = 1, \ldots, J$. Thus, we assume all locations are monitored at the same set of times. In the sequel, we adopt an equally spaced time scale, so that we are modelling a time series of spatial processes. In our development we assume that all IJ points have been observed; randomly missing values create no additional conceptual difficulty, but add computational burden. As a result, the change of support problem is only in space. At a new location $\mathbf{s}'$ we seek to predict the entire vector $(Y(\mathbf{s}', t_1), \ldots, Y(\mathbf{s}', t_J))$. Similarly, at a block B we seek to predict the vector of block averages $(Y(B, t_1), \ldots, Y(B, t_J))$.

The format of the paper is as follows. In Section 2 we develop the methodology for spatial data at a single time point, while Section 3 presents the general spatio-temporal methodology. Section 4 applies our approaches to the Atlanta ozone data. Finally, Section 5 offers some concluding remarks.

2. METHODOLOGY FOR THE STATIC SPATIAL CASE

We start with a stationary Gaussian process specification for $Y(\mathbf{s})$ having mean function $\mu(\mathbf{s}; \boldsymbol{\beta})$ and covariance function $c(\mathbf{s} - \mathbf{t}; \boldsymbol{\theta})$. Here μ is a trend surface with coefficient vector $\boldsymbol{\beta}$ while $\boldsymbol{\theta}$ denotes the parameters associated with the stationary covariance function. Beginning with point data observed at sites $\mathbf{s}_1, \ldots, \mathbf{s}_I$, let $\mathbf{Y}_s^T = (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_I))$. Then

$$\mathbf{Y}_s \mid \boldsymbol{\beta}, \boldsymbol{\theta} \sim N(\boldsymbol{\mu}_s(\boldsymbol{\beta}), H_s(\boldsymbol{\theta})) \tag{2}$$

where $\boldsymbol{\mu}_s(\boldsymbol{\beta})_i = \mu(\mathbf{s}_i; \boldsymbol{\beta})$ and $(H_s(\boldsymbol{\theta}))_{ii'} = c(\mathbf{s}_i - \mathbf{s}_{i'}; \boldsymbol{\theta})$.

Given a prior on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, models such as (2) are straightforwardly fit using simulation methods either through importance sampling (e.g. Ecker and Gelfand (1997)) or Gibbs sampling (e.g. Karson *et al.* (1999)). Regardless, we can assume the existence of posterior samples $(\boldsymbol{\beta}_g^*, \boldsymbol{\theta}_g^*)$, $g = 1, \ldots, G$ from $f(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \mathbf{Y}_s)$. The required inversion of $H_s(\boldsymbol{\theta})$ in order to evaluate the likelihood can be problematic when I is large.

Then for prediction at a set of new locations, i.e. $\mathbf{Y}_{s'}^T = (Y(\mathbf{s}'_1), \ldots, Y(\mathbf{s}'_K))$, we require only the predictive distribution

$$f(\mathbf{Y}_{s'} \mid \mathbf{Y}_s) = \int f(\mathbf{Y}_{s'} \mid \mathbf{Y}_s, \boldsymbol{\beta}, \boldsymbol{\theta}) f(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \mathbf{Y}_s) \, \mathrm{d}\boldsymbol{\beta} \, \mathrm{d}\boldsymbol{\theta}. \tag{3}$$

By drawing $\mathbf{Y}_{s',g}^* \sim f(\mathbf{Y}_{s'} \mid \mathbf{Y}_s, \boldsymbol{\beta}_g^*, \boldsymbol{\theta}_g^*)$ we obtain a sample from (3) which provides any desired inference about $\mathbf{Y}_{s'}$ and its components.

Under a Gaussian process,

$$f\left(\begin{pmatrix}\mathbf{Y}_s \\ \mathbf{Y}_{s'}\end{pmatrix} \middle| \boldsymbol{\beta}, \boldsymbol{\theta}\right) = N\left(\begin{pmatrix}\boldsymbol{\mu}_s(\boldsymbol{\beta}) \\ \boldsymbol{\mu}_{s'}(\boldsymbol{\beta})\end{pmatrix}, \begin{pmatrix}H_s(\boldsymbol{\theta}) & H_{s,s'}(\boldsymbol{\theta}) \\ H_{s,s'}^T(\boldsymbol{\theta}) & H_{s'}(\boldsymbol{\theta})\end{pmatrix}\right) \tag{4}$$

with entries defined as in (2). Hence, $\mathbf{Y}_{s'} | \mathbf{Y}_s, \boldsymbol{\beta}, \boldsymbol{\theta}$ is distributed as

$$N(\boldsymbol{\mu}_{s'}(\boldsymbol{\beta}) + H_{s,s'}^T(\boldsymbol{\theta}) H_s^{-1}(\boldsymbol{\theta})(\mathbf{Y}_s - \mu_s(\boldsymbol{\beta})), \; H_{s'}(\boldsymbol{\theta}) - H_{s,s'}^T(\boldsymbol{\theta}) H_s^{-1}(\boldsymbol{\theta}) H_{s,s'}(\boldsymbol{\theta})). \tag{5}$$

Sampling from (5) requires the inversion of $H_s(\boldsymbol{\theta}_g^*)$, which will already have been done in sampling $\boldsymbol{\theta}_g^*$, and then the square root of the $K \times K$ covariance matrix in (5).

Turning next to prediction for blocks, $B_1, \ldots, B_K$, i.e. $\mathbf{Y}_B^T = (Y(B_1), \ldots, Y(B_K))$ we again require the predictive distribution, which is now

$$f(\mathbf{Y}_B \mid \mathbf{Y}_s) = \int f(\mathbf{Y}_B \mid \mathbf{Y}_s; \boldsymbol{\beta}, \boldsymbol{\theta}) f(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \mathbf{Y}_s) \, \mathrm{d}\boldsymbol{\beta} \, \mathrm{d}\boldsymbol{\theta}. \tag{6}$$

Under a Gaussian process, we now have

$$f\left(\begin{pmatrix} \mathbf{Y}_s \\ \mathbf{Y}_B \end{pmatrix} \middle| \boldsymbol{\beta}, \boldsymbol{\theta} \right) = N\left(\begin{pmatrix} \boldsymbol{\mu}_s(\boldsymbol{\beta}) \\ \boldsymbol{\mu}_B(\boldsymbol{\beta}) \end{pmatrix}, \begin{pmatrix} H_s(\boldsymbol{\theta}) & H_{s,B}(\boldsymbol{\theta}) \\ H_{s,B}^T(\boldsymbol{\theta}) & H_B(\boldsymbol{\theta}) \end{pmatrix} \right), \tag{7}$$

where

$$\begin{aligned}
(\boldsymbol{\mu}_B(\boldsymbol{\beta}))_k &= E(Y(B_k) \mid \boldsymbol{\beta}) \;=\; |B_k|^{-1} \int_{B_k} \mu(\mathbf{s}; \boldsymbol{\beta}) \, \mathrm{d}\mathbf{s}, \\
(H_B(\boldsymbol{\theta}))_{kk'} &= |B_k|^{-1} |B_{k'}|^{-1} \int_{B_k} \int_{B_{k'}} c(\mathbf{s} - \mathbf{s}'; \boldsymbol{\theta}) \, \mathrm{d}\mathbf{s}' \, \mathrm{d}\mathbf{s}, \\
(H_{s,B}(\boldsymbol{\theta}))_{ik} &= |B_k|^{-1} \int_{B_k} c(\mathbf{s}_i - \mathbf{s}'; \boldsymbol{\theta}) \, \mathrm{d}\mathbf{s}'.
\end{aligned}$$

Analogously to (5), $\mathbf{Y}_B | \mathbf{Y}_s, \boldsymbol{\beta}, \boldsymbol{\theta}$ is distributed as

$$N(\boldsymbol{\mu}_B(\boldsymbol{\beta}) + H_{s,B}^T(\boldsymbol{\theta}) H_s^{-1}(\boldsymbol{\theta})(\mathbf{Y}_s - \boldsymbol{\mu}_s(\boldsymbol{\beta})) \;,\; H_B(\boldsymbol{\theta}) - H_{s,B}^T(\boldsymbol{\theta}) H_s^{-1}(\boldsymbol{\theta}) H_{s,B}(\boldsymbol{\theta})). \tag{8}$$

The major difference between (5) and (8) is that in (5), given $(\boldsymbol{\beta}_g^*, \boldsymbol{\theta}_g^*)$, numerical values for all of the entries in $\boldsymbol{\mu}_{s'}(\boldsymbol{\beta})$, $H_{s'}(\boldsymbol{\theta})$ and $H_{s,s'}(\boldsymbol{\theta})$ are immediately obtained. In (8) every analogous entry requires an integration as above. Anticipating irregularly shaped B_k, Riemann approximation to integrate over these regions may be awkward. Instead, noting that each such integration is an expectation with respect to a uniform distribution, we propose Monte Carlo integration. In particular, for each B_k we propose to draw a set of locations $\mathbf{s}_{k,\ell}$, $\ell = 1, 2, \ldots, L_k$, distributed independently and uniformly over B_k. Here L_k can vary with k to allow for very unequal $|B_k|$.

Hence, we replace $(\boldsymbol{\mu}_B(\boldsymbol{\beta}))_k$, $(H_B(\boldsymbol{\theta}))_{kk'}$ and $(H_{s,B}(\boldsymbol{\theta}))_{ik}$ with

$$\begin{aligned}
(\widehat{\boldsymbol{\mu}}_B(\boldsymbol{\beta}))_k &= L_k^{-1} \sum_{\ell} \mu(\mathbf{s}_{k,\ell}; \boldsymbol{\beta}), \\
(\widehat{H}_B(\boldsymbol{\theta}))_{kk'} &= L_k^{-1} L_{k'}^{-1} \sum_{\ell} \sum_{\ell'} c(\mathbf{s}_{k\ell} - \mathbf{s}_{k'\ell'}; \boldsymbol{\theta}), \\
(\widehat{H}_{s,B}(\boldsymbol{\theta}))_{ik} &= L_k^{-1} \sum_{\ell} c(\mathbf{s}_i - \mathbf{s}_{k\ell}; \boldsymbol{\theta}).
\end{aligned} \tag{9}$$

In our notation, the 'hat' denotes a Monte Carlo integration which can be made arbitrarily accurate and has nothing to do with the data $\mathbf{Y}_s$. Note also that the same set of $\mathbf{s}_{k\ell}$ can be used for each integration and with each $(\boldsymbol{\beta}_g^*, \boldsymbol{\theta}_g^*)$; we need only obtain this set once. In obvious notation we replace (7) with the $(I + K)$-dimensional multivariate normal distribution $\widehat{f}((\mathbf{Y}_s, \mathbf{Y}_B)^T \mid \boldsymbol{\beta}, \boldsymbol{\theta})$.

It is useful to note that, if we define $\widehat{Y}(B_k) = L_k^{-1} \sum_{\ell} Y(\mathbf{s}_{k\ell})$, then $\widehat{Y}(B_k)$ is a Monte Carlo integration for $Y(B_k)$ as given in (1). Then, with an obvious definition for $\widehat{\mathbf{Y}}_B$, it is apparent that

$$\widehat{f}((\mathbf{Y}_s, \mathbf{Y}_B)^T \mid \boldsymbol{\beta}, \boldsymbol{\theta}) = f((\mathbf{Y}_s, \widehat{\mathbf{Y}}_B)^T \mid \boldsymbol{\beta}, \boldsymbol{\theta}) \tag{10}$$

where (10) is interpreted to mean that the approximate joint distribution of $(\mathbf{Y}_s, \mathbf{Y}_B)$ is the exact joint distribution of $\mathbf{Y}_s, \widehat{\mathbf{Y}}_B$. In practice, we will work with $\widehat{f}$, converting to $\widehat{f}(\mathbf{Y}_B \mid \mathbf{Y}_s, \boldsymbol{\beta}, \boldsymbol{\theta})$ to sample $\mathbf{Y}_B$ rather than sampling the $\widehat{Y}(B_k)$ through the $Y(\mathbf{s}_{k\ell})$. But, evidently, we are sampling $\widehat{\mathbf{Y}}_B$ rather than $\mathbf{Y}_B$.

Hence, we ask when $\widehat{\mathbf{Y}}_B \xrightarrow{P} \mathbf{Y}_B$. An obvious sufficient condition is that realizations of the $Y(\mathbf{s})$ process are almost surely continuous. In the stationary case, Kent (1989) provides sufficient conditions on $c(\mathbf{s}-\mathbf{t};\boldsymbol{\theta})$ to ensure this. Alternatively, Stein (1999) defines $Y(\mathbf{s})$ to be *mean square continuous* if $\lim_{\mathbf{h}\to 0} E(Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s}))^2 = 0$ for all $\mathbf{s}$. But then $Y(\mathbf{s}+\mathbf{h}) \xrightarrow{P} Y(\mathbf{s})$ as $\mathbf{h} \to 0$, which is sufficient to guarantee that $\widehat{\mathbf{Y}}_B \xrightarrow{P} \mathbf{Y}_B$. Stein notes that if $Y(s)$ is stationary, we only require $c(\cdot\,;\,\boldsymbol{\theta})$ continuous at $\mathbf{0}$ for mean square continuity.

Finally, suppose we start with block data, i.e. we observe $\mathbf{Y}_B^T = (Y(B_1), \ldots, Y(B_I))$. Then, analogous to (2), the likelihood is well defined, i.e.

$$f(\mathbf{Y}_B \mid \boldsymbol{\beta}, \boldsymbol{\theta}) = N(\boldsymbol{\mu}_B(\boldsymbol{\beta}), H_B(\boldsymbol{\theta})). \tag{11}$$

Hence, given a prior on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, the Bayesian model is completely specified. As above, evaluation of the likelihood requires integrations. So, we replace (11) with

$$\widehat{f}(\mathbf{Y}_B \mid \boldsymbol{\beta}, \boldsymbol{\theta}) = N(\widehat{\boldsymbol{\mu}}_B(\boldsymbol{\beta}), \widehat{H}_B(\boldsymbol{\theta})). \tag{12}$$

Simulation-based fitting is now straightforward, as below (2), albeit somewhat more time-consuming due to the need to calculate $\widehat{\boldsymbol{\mu}}_B(\boldsymbol{\beta})$ and $\widehat{H}_B(\boldsymbol{\theta})$.

To predict for $\mathbf{Y}_{s'}$ we require $f(\mathbf{Y}_{s'} \mid \mathbf{Y}_B)$. As above, we only require $f(\mathbf{Y}_B, \mathbf{Y}_{s'} \mid \boldsymbol{\beta}, \boldsymbol{\theta})$ which has been given in (7). Using (10) we now obtain $\widehat{f}(\mathbf{Y}_{s'} \mid \mathbf{Y}_B, \boldsymbol{\beta}, \boldsymbol{\theta})$ to sample $\mathbf{Y}_{s'}$. Note that $\widehat{f}$ is used in (12) to obtain the posterior samples and again to obtain the predictive samples. Equivalently, the foregoing discussion shows that we can replace $\mathbf{Y}_B$ with $\widehat{\mathbf{Y}}_B$ throughout. To predict for new blocks $B'_1, \ldots, B'_K$, let $\mathbf{Y}_{B'}^T = (Y(B'_1), \ldots, Y(B'_K))$. Now we require $f(\mathbf{Y}_{B'} \mid \mathbf{Y}_B)$, which in turn requires $f(\mathbf{Y}_B, \mathbf{Y}_{B'} \mid \boldsymbol{\beta}, \boldsymbol{\theta})$. The approximate distribution $\widehat{f}(\mathbf{Y}_B, \mathbf{Y}_{B'} | \boldsymbol{\beta}, \boldsymbol{\theta})$ employs Monte Carlo integrations over the B'_k as well as the B_i, and yields $\widehat{f}(\mathbf{Y}_{B'} \mid \mathbf{Y}_B, \boldsymbol{\beta}, \boldsymbol{\theta})$ to sample $\mathbf{Y}_{B'}$. Again $\widehat{f}$ is used to obtain both the posterior and predictive samples.

Note that, in all four prediction cases, we can confine ourselves to an $(I+K)$-dimensional multivariate normal. Moreover, we have only an $I \times I$ matrix to invert repeatedly in the model fitting, and a $K \times K$ matrix whose square root is required for the predictive sampling.

For the modifiable areal unit problem (i.e. prediction at new blocks using data for a given set of blocks), suppose we take as our point estimate for a generic new set B_0 the posterior mean

$$E(Y(B_0) \mid \mathbf{Y}_B) = E\{\mu(B_0;\boldsymbol{\beta}) + \mathbf{H}_{B,B_0}^T(\boldsymbol{\theta}) H_B^{-1}(\boldsymbol{\theta})(\mathbf{Y}_B - \boldsymbol{\mu}_B(\boldsymbol{\beta})) \mid \mathbf{Y}_B\},$$

where $\mathbf{H}_{B,B_0}(\boldsymbol{\theta})$ is $I \times 1$ with ith entry equal to $\operatorname{cov}(Y(B_i), Y(B_0) \mid \boldsymbol{\theta})$. If $\mu(\mathbf{s};\boldsymbol{\beta}) \equiv \mu_i$ for $\mathbf{s} \in B_i$, then $\mu(B_0;\boldsymbol{\beta}) = |B_0|^{-1} \sum_i |B_i \cap B_0| \mu_i$. But $E(\mu_i \mid \mathbf{Y}_B) \approx Y(B_i)$ to a first-order approximation, so in this case $E(Y(B_0) \mid \mathbf{Y}_B) \approx |B_0|^{-1} \sum_i |B_i \cap B_0| Y(B_i)$, the areally weighted estimate.

3. METHODOLOGY FOR SPATIO-TEMPORAL DATA

Following Section 2, we start with point-referenced data, $Y(\mathbf{s}_i, t_j)$, $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, J$ and specify a stationary Gaussian multivariate process for $\mathbf{Y}(\mathbf{s})^T = (Y(\mathbf{s}, t_1), \ldots, Y(\mathbf{s}, t_J))$. In particular, we set $E(Y(\mathbf{s}, t) \mid \boldsymbol{\beta}) = \boldsymbol{\mu}(\mathbf{s}, t; \boldsymbol{\beta})$, a temporal trend surface model (i.e. a polynomial in $\mathbf{s}$ and t). As for the cross-covariance structure, i.e. the matrix of covariances between $\mathbf{Y}(\mathbf{s}_i)$ and $\mathbf{Y}(\mathbf{s}_{i'})$, we assume the (j, j') entry,

$$\operatorname{cov}(Y(\mathbf{s}_i, t_j), Y(\mathbf{s}_{i'}, t_{j'})) = \sigma^2 c^{(1)}(\mathbf{s}_i - \mathbf{s}_{i'}; \phi) \cdot c^{(2)}(t_j - t_{j'}; \boldsymbol{\rho}), \tag{13}$$

where $c^{(1)}$ is a valid two-dimensional correlation function and $c^{(2)}$ is a valid one-dimensional correlation function. The multiplicative form conveniently separates space and time in the calculations below, enabling feasible computation for the change of support problem. In view of the discreteness of the time scale, (13) suggests that we may view the entire specification as a time series of spatial processes. Spatial association at a fixed time point is captured through $c^{(1)}$; decay in such association over time is captured by $c^{(2)}$. Forms such as (13) have a history in spatio-temporal modelling: see, for example, Mardia and Goodall (1993) and references therein.

Collecting all the observed data into a vector $\mathbf{Y}_s^T = (\mathbf{Y}^T(\mathbf{s}_1), \ldots, \mathbf{Y}^T(\mathbf{s}_I))$ we have $\mathbf{Y}_s$ distributed as an IJ-dimensional multivariate normal with, in obvious notation, mean vector $\boldsymbol{\mu}_s(\boldsymbol{\beta})$ and covariance matrix

$$\Sigma_{\mathbf{Y}_s}(\sigma^2, \boldsymbol{\phi}, \boldsymbol{\rho}) = \sigma^2 H_s(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho}), \tag{14}$$

where '$\otimes$' denotes the Kronecker product. In (14), $H_s(\boldsymbol{\phi})$ is $I \times I$ with $(H_s(\boldsymbol{\phi}))_{ii'} = c^{(1)}(\mathbf{s}_i - \mathbf{s}_i'; \boldsymbol{\theta})$, and $H_t(\boldsymbol{\rho})$ is $J \times J$ with $(H_t(\boldsymbol{\rho}))_{jj'} = c^{(2)}(t_j - t_{j'}; \boldsymbol{\rho})$.

Given a prior for $\boldsymbol{\beta}$, σ^2, $\boldsymbol{\phi}$, and $\boldsymbol{\rho}$, the Bayesian model is completely specified. Simulation-based model fitting can be carried out similarly to the static spatial case by noting the following. The log likelihood arising from $\mathbf{Y}_s$ is

$$-\frac{1}{2}\log|\sigma^2 H_s(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho})| - \frac{1}{2\sigma^2}(\mathbf{Y}_s - \boldsymbol{\mu}_s(\boldsymbol{\beta}))^T (H_s(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho}))^{-1}(\mathbf{Y}_s - \boldsymbol{\mu}_s(\boldsymbol{\beta})).$$

But $|\sigma^2 H_s(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho})| = (\sigma^2)^{IJ}|H_s(\boldsymbol{\phi})|^J |H_t(\boldsymbol{\rho})|^I$, and $(H_s(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho}))^{-1} = H_s^{-1}(\boldsymbol{\phi}) \otimes H_t^{-1}(\boldsymbol{\rho})$. In other words, even though (14) is $IJ \times IJ$, we need only the determinant and inverse for an $I \times I$ and a $J \times J$ matrix, so that Gibbs sampling is tractable.

With regard to prediction, first consider new locations $\mathbf{s}_1', \ldots, \mathbf{s}_k'$ with interest in inference for $Y(\mathbf{s}_k', t_j)$. As with the observed data, we collect the $Y(\mathbf{s}_k', t_j)$ into vectors $\mathbf{Y}(\mathbf{s}_k')$, and the $\mathbf{Y}(\mathbf{s}_k')$ into a single $KJ \times 1$ vector $\mathbf{Y}_{s'}$. Even though we may not necessarily be interested in every component of $\mathbf{Y}_{s'}$, the simplifying forms which follow below suggest that, with regard to programming, it may be easiest to simulate draws from the entire predictive distribution $f(\mathbf{Y}_{s'} \mid \mathbf{Y}_s)$ and then retain only the desired components.

Since $f(\mathbf{Y}_{s'}|\mathbf{Y}_s)$ has a form analogous to (3), given posterior samples $(\boldsymbol{\beta}_g^*, \sigma_g^{2*}, \boldsymbol{\phi}_g^*, \boldsymbol{\rho}_g^*)$, $g = 1, \ldots, G$, we draw $\mathbf{Y}_{s',g}^*$ from $f(\mathbf{Y}_{s'} \mid \mathbf{Y}_s, \boldsymbol{\beta}_g^*, \sigma_g^{2*}, \boldsymbol{\phi}_g^*, \boldsymbol{\rho}_g^*)$. Analogous to (4),

$$f\left(\begin{pmatrix}\mathbf{Y}_s \\ \mathbf{Y}_{s'}\end{pmatrix}\middle| \boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\rho}\right) = N\left(\begin{pmatrix}\boldsymbol{\mu}_s(\boldsymbol{\beta}) \\ \boldsymbol{\mu}_{s'}(\boldsymbol{\beta})\end{pmatrix}, \sigma^2 \begin{pmatrix}H_s(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho}) & H_{s,s'}(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho}) \\ H_{s,s'}^T(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho}) & H_{s'}(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho})\end{pmatrix}\right), \tag{15}$$

with obvious definitions for $H_{s'}(\boldsymbol{\phi})$ and $H_{s,s'}(\boldsymbol{\phi})$. But then $\mathbf{Y}_{s'} \mid \mathbf{Y}_s, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\rho}$ is also normally distributed, with mean

$$\begin{aligned}\boldsymbol{\mu}_{s'}(\boldsymbol{\beta}) + (H_{s,s'}^T(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\phi}))(H_s(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho}))^{-1}(Y_s - \boldsymbol{\mu}_s(\boldsymbol{\beta})) \\ = \boldsymbol{\mu}_{s'}(\boldsymbol{\beta}) + (H_{s,s'}^T(\boldsymbol{\phi})H_s^{-1}(\boldsymbol{\phi}) \otimes I_{J\times J})(\mathbf{Y}_s - \boldsymbol{\mu}_s(\boldsymbol{\beta})),\end{aligned} \tag{16}$$

and variance

$$\begin{aligned}H_{s'}(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho}) - (H_{s,s'}^T \otimes H_t(\boldsymbol{\rho}))(H_s(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho}))^{-1}(H_{s,s'}(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho})) \\ = (H_{s'}(\boldsymbol{\phi}) - H_{s,s'}^T(\boldsymbol{\phi})H_s^{-1}(\boldsymbol{\phi})H_{s,s}(\boldsymbol{\phi})) \otimes H_t(\boldsymbol{\rho}),\end{aligned} \tag{17}$$

using standard properties of Kronecker products. In (16), time disappears apart from $\boldsymbol{\mu}_{s'}(\boldsymbol{\beta})$, while in (17), time 'factors out' of the conditioning. Sampling from this normal distribution does require the inverse square root of the conditional covariance matrix, but conveniently, this is

$$(H_{s'}(\boldsymbol{\phi}) - H_{s,s'}^T(\boldsymbol{\phi})H_s^{-1}(\boldsymbol{\phi})H_{s,s'}(\boldsymbol{\phi}))^{-\frac{1}{2}} \otimes H_t^{-\frac{1}{2}}(\boldsymbol{\rho}),$$

so the only work required beyond that in (5) is obtaining $H_t^{-\frac{1}{2}}(\boldsymbol{\rho})$, since $H_t^{-1}(\boldsymbol{\rho})$ will already have been obtained in evaluating the likelihood, following the discussion above.

Returning to blocks $B_1, \ldots, B_K$, we set $\mathbf{Y}^T(B_k) = (Y(B_k, t_1), \ldots, Y(B_k, t_J))$ and then set $\mathbf{Y}_B^T = (\mathbf{Y}^T(B_1), \ldots, \mathbf{Y}^T(B_K))$. Analogous to (6) we seek to sample $f(\mathbf{Y}_B \mid \mathbf{Y}_s)$, so we require $f(\mathbf{Y}_B \mid \mathbf{Y}_s, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\rho})$. Analogous to (15), this is

$$f\left(\begin{pmatrix}\mathbf{Y}_s\\ \mathbf{Y}_B\end{pmatrix}\middle| \boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\rho}\right) = N\left(\begin{pmatrix}\boldsymbol{\mu}_s(\boldsymbol{\beta})\\ \boldsymbol{\mu}_B(\boldsymbol{\beta})\end{pmatrix}, \sigma^2\begin{pmatrix}H_s(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho}) & H_{s,B}(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho})\\ H_{s,B}^T(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho}) & H_B(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho})\end{pmatrix}\right),$$

with $\boldsymbol{\mu}_B(\boldsymbol{\beta})$, $H_B(\boldsymbol{\phi})$, and $H_{s,B}(\boldsymbol{\phi})$ defined as in Section 2. Hence, $f(\mathbf{Y}\mathbf{Y}_s, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\rho})$ is again normal with mean and variance as given in (16) and (17), but with $\boldsymbol{\mu}_B(\boldsymbol{\beta})$ replacing $\boldsymbol{\mu}_{s'}(\boldsymbol{\beta})$, $H_B(\boldsymbol{\phi})$ replacing $H_{s'}(\boldsymbol{\phi})$, and $H_{s,B}(\boldsymbol{\phi})$ replacing $H_{s,s'}(\boldsymbol{\phi})$. Using the same Monte Carlo integrations as proposed in Section 2 leads to sampling the resultant $\widehat{f}(\mathbf{Y}_B \mid \mathbf{Y}_s, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\rho})$, and the same technical justification applies.

If we started with block data, $Y(B_i, t_j)$, then following (11) and (14),

$$f(\mathbf{Y}_B \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\rho}) = N(\mu_B(\boldsymbol{\beta}), \sigma^2(H_B(\boldsymbol{\phi}) \otimes H_t(\boldsymbol{\rho})). \tag{18}$$

Given (18), the path for prediction at new points or at new blocks is clear, following the above and the end of Section 2; we omit the details.

Finally, though our interest is in change of support at observed times, the association structure in (13) allows *forecasting* of the spatial process at time t_{J+1} in the series. This can be done at observed or unobserved points or blocks following the foregoing development. To retain the above simplifying forms, we would first simulate the variables at t_{J+1} associated with observed points or blocks (with no change of support). We would then revise $H_t(\boldsymbol{\phi})$ to be $(J+1) \times (J+1)$ before proceeding as above.

4. ANALYSIS OF ATLANTA OZONE DATA

4.1 *Static spatial case*

In this section, we use the approach of Section 2 to perform point–point and point–block inference for the Atlanta ozone data pictured in Figure 1. Recall that the target points are those marked A and B on the map, while the target blocks are the 36 Atlanta city zips. The differing block sizes suggest use of a different L_k for each k in equation (9). Conveniently, our GIS, `ARC/INFO`, can generate random points over the whole study area, and then allocate them to each zip. Thus L_k is proportional to the area of the zip, $|B_k|$. Our procedure produced 3743 randomly chosen locations distributed over the 36 city zips, for an average L_k of nearly 104.

Suppose that log-ozone exposure $Y(\mathbf{s})$ follows a second-order stationary spatial Gaussian process, using the simple spatial covariance function $c(\mathbf{s}_i - \mathbf{s}_{i'}; \boldsymbol{\theta}) = \sigma^2 \mathrm{e}^{-\phi\|\mathbf{s}_i - \mathbf{s}_{i'}\|}$, where $\|\mathbf{s}_i - \mathbf{s}_{i'}\|$ is the Euclidean distance between sites $\mathbf{s}_i$ and $\mathbf{s}_{i'}$. A preliminary exploratory analysis of our dataset suggested a constant mean function $\mu(\mathbf{s}_i; \boldsymbol{\beta}) = \mu$ is adequate for our dataset: cf. Figure 3 and the associated discussion in Section 4.2 below. We place the customary flat prior on μ, and assume that $\sigma^2 \sim IG(a, b)$ and

$\phi \sim G(c, d)$ where IG and G denote the inverse gamma and gamma distributions with probability density functions $p(\sigma^2|a, b) = \frac{e^{-1/(b\sigma^2)}}{\Gamma(a)b^a(\sigma^2)^{a+1}}$ and $p(\phi|c, d) = \frac{\phi^{c-1}e^{-\phi/d}}{\Gamma(c)d^c}$ respectively. We chose $a = 3$, $b = 0.5$, $c = 0.03$ and $d = 100$, corresponding to fairly vague priors. We then fit this three-parameter model using an MCMC implementation, which ran three parallel sampling chains for 1000 iterations each, sampling μ and σ^2 via Gibbs steps and ϕ through Metropolis–Hastings steps with a $G(3, 1)$ candidate density. Convergence of the sampling chains was virtually immediate. We obtained the following posterior medians and 95% equal-tail credible intervals for the three parameters: for μ, 0.111 and (0.072, 0.167); for σ^2, 1.37 and (1.18, 2.11); and for ϕ, 1.62 and (0.28, 0.13).

Figure 2 maps summaries of the posterior samples for the 36 target blocks (city zips) and the two target points (A and B); specifically, the posterior medians, $q_{0.50}$, upper and lower 0.025 points, $q_{0.975}$ and $q_{0.025}$, and the lengths of the 95% equal-tail credible intervals, $q_{0.975} - q_{0.025}$. The zip-level medians show a clear spatial pattern, with the highest predicted block averages occurring in the southeastern part of the city near the two high observed readings (0.144 and 0.136), and the lower predictions in the north apparently the result of smoothing toward the low observed value in this direction (0.076). The interval lengths reflect spatial variability, with lower values occurring in larger areas (which require more averaging) or in areas nearer to observed monitoring stations (e.g. those near the southeastern, northeastern, and western city boundaries). Finally, note that our approach allows sensibly differing predicted medians for points A and B, with A being higher due to the slope of the fitted surface. Previous centroid-based analyses (like that of Carlin *et al.* (1999)) would instead implausibly impute the same fitted value to both points, since both lie within the same zip.

4.2 *Spatio-temporal case*

To illustrate the method of Section 3, we use a spatio-temporal version of the Atlanta ozone dataset. As mentioned in Section 1, we actually have ozone measurements at the 10 fixed monitoring stations shown in Figure 1 over the 92 summer days in 1995. Figure 3 shows the daily one-hour maximum ozone reading for the sites during July of this year. There are several sharp peaks, but little evidence of a weekly (7 day) period in the data. The mean structure appears reasonably constant in space, with the ordering of the site measurements changing dramatically for different days. Moreover, with only 10 'design points' in the metro area, any spatial trend surface we fit would be quite speculative over much of the study region (e.g. the northwest and southwest metro; see Figure 1). The temporal evolution of the series is not inconsistent with a constant mean autoregressive error model; indeed, the lag 1 sample autocorrelation varies between 0.27 and 0.73 over the 10 sites, strongly suggesting the need for a model accounting for both spatial and temporal correlations.

We thus fit our spatio-temporal model with mean $\boldsymbol{\mu}(\mathbf{s}, t; \boldsymbol{\beta}) = \mu$, but with spatial and temporal correlation functions $c^{(1)}(\mathbf{s}_i - \mathbf{s}_{i'}; \phi) = e^{-\phi\|\mathbf{s}_i - \mathbf{s}_{i'}\|}$ and $c^{(2)}(t_j - t_{j'}; \rho) = \rho^{|j-j'|}/(1 - \rho^2)$. Hence our model has four parameters: we use a flat prior for μ, an $IG(3, 0.5)$ prior for σ^2, a $G(0.003, 100)$ prior for ϕ, and a $U(0, 1)$ prior for ρ (thus eliminating the implausible possibility of *negative* autocorrelation in our data, but favouring no positive value over any other). To facilitate our Gibbs–Metropolis approach, we transform to $\theta = \log\phi$ and $\lambda = \log(\rho/(1 - \rho))$, and subsequently use Gaussian proposals on these transformed parameters.

Running three parallel chains of 10 000 iterations each, sample traces (not shown) again indicate virtually immediate convergence of our algorithm. Posterior medians and 95% equal-tail credible intervals for the four parameters are as follows: for μ, 0.068 and (0.057, 0.080); for σ^2, 0.11 and (0.08, 0.17); for ϕ, 0.06 and (0.03, 0.08); and for ρ, 0.42 and (0.31, 0.52). The rather large value of ρ confirms the strong temporal autocorrelation suspected in the daily ozone readings.

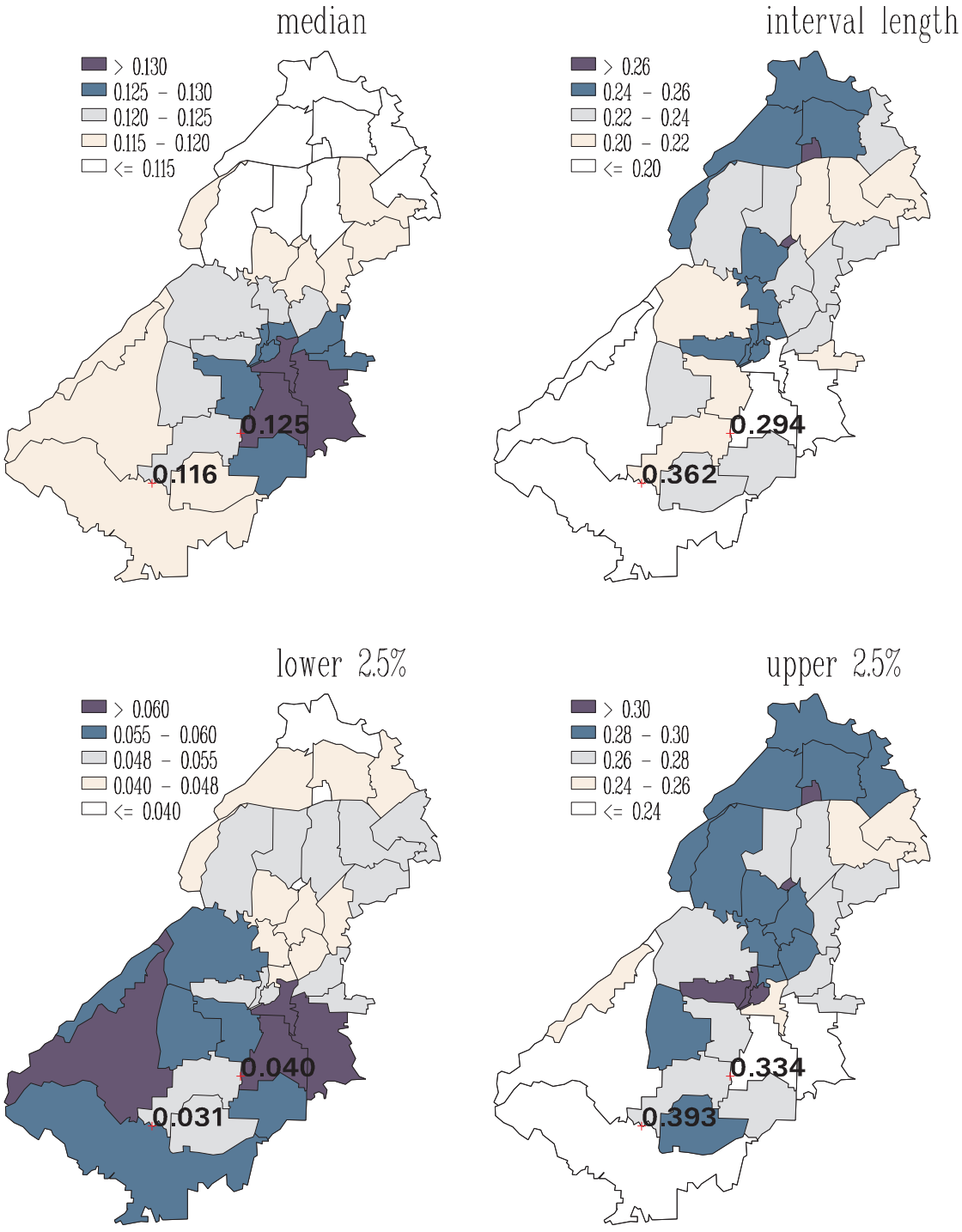

Fig. 2. Posterior point–point and point–block summaries, static spatial model, Atlanta ozone data for July 15, 1995.

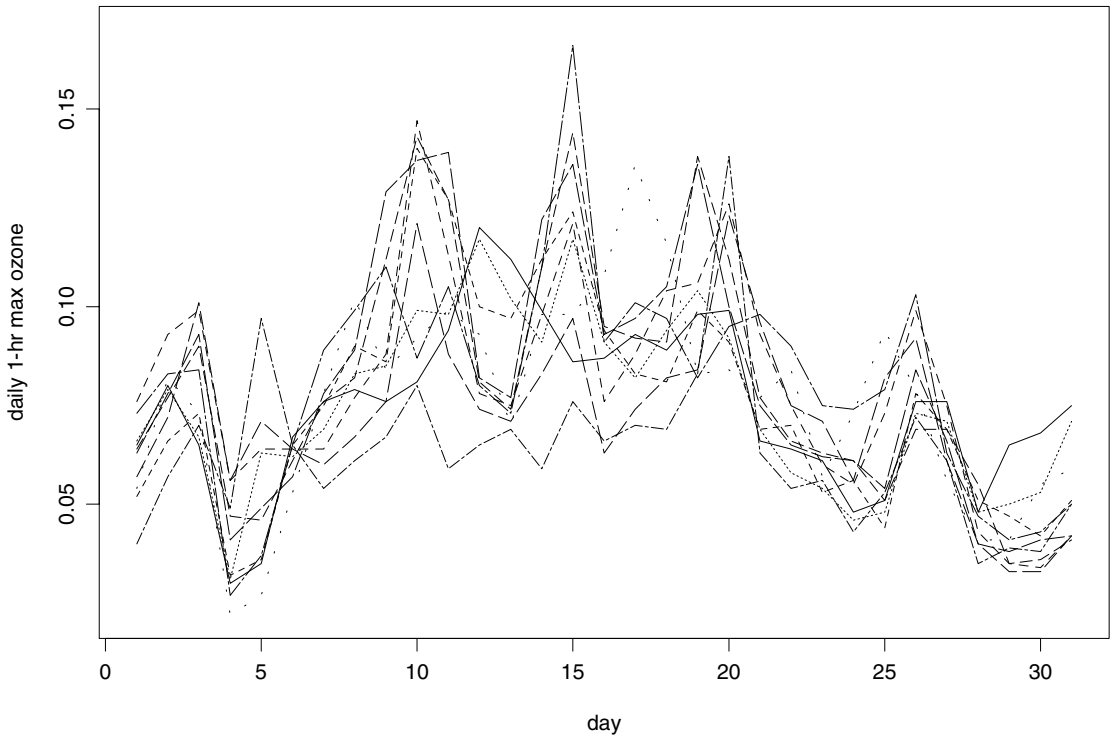

Fig. 3. One-hour maximum ozone by day, July 1995, 10 Atlanta monitoring sites.

Comparison of the posteriors for σ^2 and ϕ with those obtained for the static spatial model in Section 4.1 is not apt, since these parameters have different meanings in the two models. Instead, we make this comparison in the context of point–point and point–block prediction. Table 1 provides posterior predictive summaries for the ozone concentrations for July 15, 1995 at points A and B (see Figure 1), as well as for the block averages over three selected Atlanta city zips: 30317, an east-central city zip very near to two monitoring sites; 30344, the south-central zip containing the points A and B; and 30350, the northernmost city zip. Results are shown for both the spatio-temporal model of this section and for the static spatial model previously fit in Section 4.1. Note that all the posterior medians are a bit higher under the spatio-temporal model, except for that for the northern zip, which remains low. Also note the significant increase in precision afforded by this model, which makes use of the data from all 31 days in July, 1995, instead of only that from July 15. Figure 4 shows the estimated posteriors giving rise to the first and last rows in Table 1 (i.e. corresponding to the July 15, 1995 ozone levels at point A and the block average over the northernmost city zip, 30350). The Bayesian approach's ability to reflect differing amounts of predictive uncertainty for the two models is clearly evident.

Finally, Figure 5 plots the posterior medians and upper and lower 0.025 quantiles produced by the spatio-temporal model by day for the ozone concentration at point A, as well as those for the block average in zip 30350. Note that the overall temporal pattern is quite similar to that for the data shown in Figure 3. Since point A is rather nearer to several data observation points, the confidence bands associated with it are often a bit narrower than those for the northern zip, but this pattern is not perfectly consistent over time. Also note that the relative positions of the bands for July 15 are consistent with the data pattern for this day

Table 1. *Posterior medians and* 95% *equal-tail credible intervals for ozone levels at two points, and for average ozone levels over three blocks (zip codes), purely spatial model versus spatio-temporal model, Atlanta ozone data for July* 15*,* 1995

	Spatial only		Spatio-temporal	
	Point	95% interval	Point	95% interval
Point A	0.125	(0.040, 0.334)	0.139	(0.111, 0.169)
Point B	0.116	(0.031, 0.393)	0.131	(0.098, 0.169)
Zip 30317 (east-central)	0.130	(0.055, 0.270)	0.138	(0.121, 0.155)
Zip 30344 (south-central)	0.123	(0.055, 0.270)	0.135	(0.112, 0.161)
Zip 30350 (north)	0.112	(0.040, 0.283)	0.109	(0.084, 0.140)

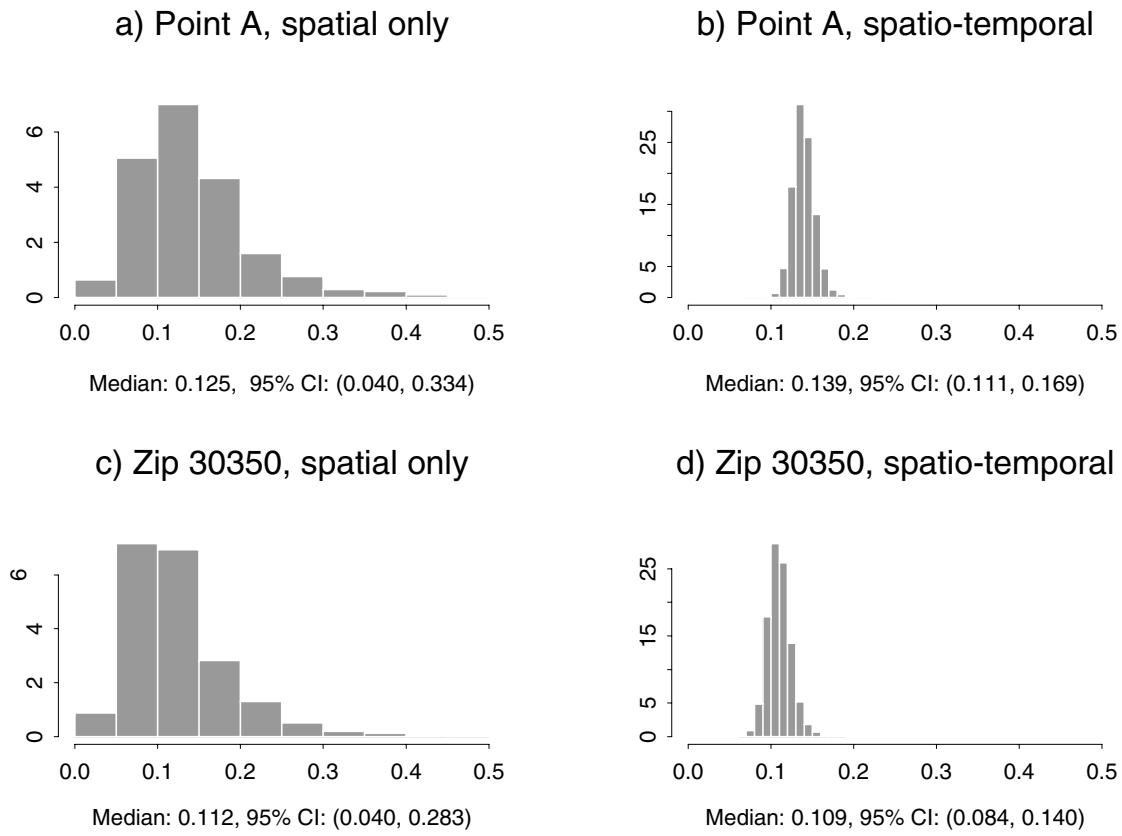

Fig. 4. Posterior predictive distributions for ozone concentration at point A and the block average over zip 30350, purely spatial model versus spatio-temporal model, Atlanta ozone data for July 15, 1995.

seen in Figure 1, when downtown exposures were higher than those in the northern metro. Finally, the day-to-day variability in the predicted series is substantially larger than the predictive variability associated with any given day.

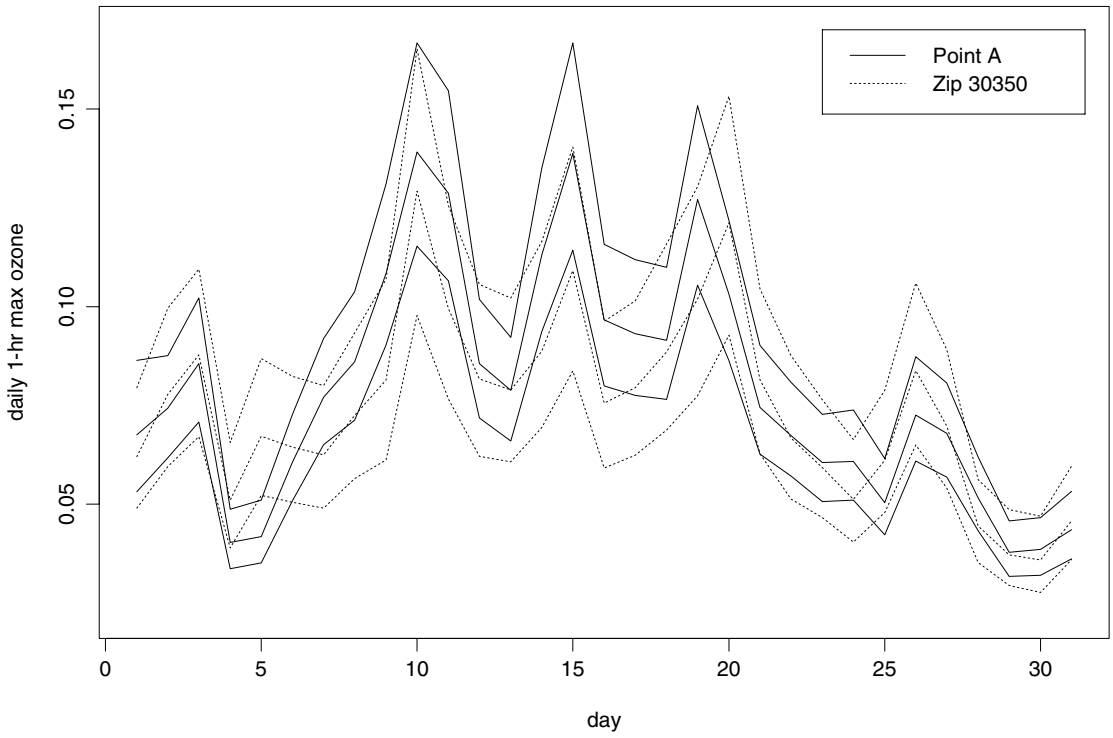

Fig. 5. Posterior medians and upper and lower 0.025 quantiles for the predicted one-hour maximum ozone concentration by day, July 1995; solid lines, point A; dotted lines, block average over zip 30350 (northernmost Atlanta city zip).

5. Concluding remarks and future directions

With continuous measurements on $\mathbb{R}^1$ (directly or by suitable transformation) and without replication, there is little hope of distinguishing a joint Gaussian distribution for the data from a heavier-tailed choice, such as a t with low degrees of freedom v. To clarify, if $\mathbf{W} \equiv (W(\mathbf{s}_1), \ldots, W(\mathbf{s}_n))$ has mean $\mathbf{0}$ and covariance matrix $\sigma^2\Sigma_W$, then $\mathbf{Z} = \Sigma_W^{-\frac{1}{2}}\mathbf{W}$ is distributed as $N(0, \sigma^2 I)$ under the normal model, but $t_v(0, \sigma^2 I)$ under the t model. Though the components of $\mathbf{Z}$ are independent under the normal and dependent under the t, in either case they are identically distributed and uncorrelated. It will thus be difficult to effectively diagnose non-normality.

Our data analyses were primarily intended as illustrative, using simple exponential forms for the covariance function $c(\mathbf{s}_i - \mathbf{s}_{i'}; \boldsymbol{\theta})$ in Section 4.1 and $c^{(1)}(\mathbf{s}_i - \mathbf{s}_{i'}; \phi)$ in Section 4.2. However, other isotropic and geometrically anisotropic forms could be fitted just as easily using our approach, since they lead to only small increases in the dimension of the parameter space and the associated computational burden. Taking this idea a step further, one might wish to shed the stationarity assumption entirely. Here a rich class of nonstationary models can be created through forms $Y(\mathbf{s}) = \sum_i a_i(\mathbf{s})Y_i(\mathbf{s})$ where the Y_i are independent stationary processes with covariance functions c_i and the $a_i(\mathbf{s})$ are specified functions (Fuentes, 2000). In practice c_i would have associated parameters $\boldsymbol{\theta}_i$, whence

$$\operatorname{cov}(Y(\mathbf{s}), Y(\mathbf{s}')) = \sum_i a_i(\mathbf{s})a_i(\mathbf{s}')c_i(\mathbf{s} - \mathbf{s}'; \boldsymbol{\theta}_i). \tag{19}$$

Hence in Section 2 (and therefore, implicitly, in Section 3), we can replace $c(\mathbf{s} - \mathbf{s}'; \boldsymbol{\theta})$ with (19).

Lastly, for point-referenced data which are not continuous (e.g. categorical or count data), only prediction at new locations is meaningful. Bayesian kriging in such cases is by now well discussed. An approach for indicator kriging using clipped Gaussian fields is discussed in DeOliveira (2000); a more general approach using hierarchical models is developed in Diggle *et al.* (1998).

ACKNOWLEDGEMENTS

The work of the first author was supported in part by NSF grant DMS 99-71206; that of the first and second authors was supported in part by NSF/EPA grant SES 99-78238; and that of the second and third authors was supported in part by National Institute of Environmental Health Sciences (NIEHS) Grant 1-R01-ES07750. The authors are grateful to Professor Paige Tolbert and Professor James Mulholland for providing and permitting analysis of the Atlanta ozone dataset.

REFERENCES

ABRAHAMSEN, N. (1993). Bayesian kriging for seismic depth conversion of a multi-layer reservoir. In Soares, A. (ed.), *Geostatistics Troia, '92*, Boston: Kluwer, pp. 385–398.

BROWN, P., LE, N. AND ZIDEK, J. (1994). Multivariate spatial interpolation and exposure to air pollutants. *The Canadian Journal of Statistics* **22**, 489–509.

CARLIN, B. P., XIA, H., DEVINE, O., TOLBERT, P. AND MULHOLLAND, J. (1999). Spatio-temporal hierarchical models for analysing Atlanta paediatric asthma ER visit rates. In Gatsonis *et al.*, C. (ed.), *Case Studies in Bayesian Statistics*, vol IV. New York: Springer, pp. 303–320.

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

CRESSIE, N. A.C. (1996). Change of support and the modifiable areal unit problem. *Geographical Systems* **3**, 159–180.

DEOLIVEIRA, V. (2000). Bayesian prediction of clipped Gaussian random fields. *Computational Statistics and Data Analysis* **34**, 299–314.

DEOLIVEIRA, V., KEDEM, B. AND SHORT, D. A. (1997). Bayesian prediction of transformed Gaussian random fields. *Journal of the American Statistical Association.* **92**, 1422–1433.

DIGGLE, P. J., TAWN, J. A. AND MOYEED, R. A. (1998). Model-based geostatistics (with discussion). *Journal of the Royal Statistical Society, Series* C *(Applied Statistics)* **47**, 299–350.

ECKER, M. D. AND GELFAND, A. E. (1997). Bayesian variogram modelling for an isotropic spatial process. *Journal of Agricultural Biological and Environmental Statistics* **2**, 347–369.

FLOWERDEW, R. AND GREEN, M. (1992). Developments in areal interpolating methods and GIS. *Annals of Regional Science* **26**, 67–78.

FUENTES, M. (2000). A new high frequency kriging approach for nonstationary environmental processes. *Technical Report*, Department of Statistics, North Carolina State University.

HANDCOCK, M. S. AND STEIN, M. L. (1993). A Bayesian analysis of kriging. *Technometrics* **35**, 403–410.

HANDCOCK, M. S. AND WALLIS, J. (1994). An approach to statistical spatial-temporal modelling of meteorological fields (with discussion). *Journal of the American Statistical Association* **89**, 368–390.

HJORT, N. AND OMRE, H. (1994). Topics in spatial statistics (with discussion). *Scandinevian Journal of Statistics* **21**, 289–357.

KARSON, M. J., GAUDARD, M., LINDER, E. AND SINHA, D. (1999). Bayesian analysis and computations for spatial prediction (with discussion). *Environmental and Ecological Statistics* **6**, 147–182.

KENT, J. T. (1989). Continuity properties for random fields. *Annals of Prob.* **17**, 1432–1440.

LE, N. AND ZIDEK, J. (1992). Interpolation with uncertain spatial covariances: a Bayesian alternative to kriging. *J. Mult. Anal.* **43**, 351–374.

MARDIA, K. V. AND GOODALL, C. (1993). Spatio-temporal analyses of multivariate environmental monitoring data. In Patil, G. P. and Rao, C. R. (eds), *Multivariate Environmental Statistics*, Amsterdam: Elsevier, pp. 347–386.

MUGGLIN, A. S. AND CARLIN, B. P. (1998). Hierarchical modelling in geographic information systems: population interpolation over incompatible zones. *Journal of Agricultural, Biological, and Environmental Statistics* **3**, 111–130.

MUGGLIN, A. S,, CARLIN, B. P. AND GELFAND, A. E. (2000). Fully model-based approaches for spatially misaligned data. *Journal of the American Statistical Association* **95**, 877–887.

OMRE, H. (1987). Bayesian kriging—merging observations and qualified guesses in kriging. *Math. Geology* **19**, 25–39.

OMRE, H. (1988). A Bayesian approach to surface estimation. In Chung *et al.*, C. F. (ed.), *Quantitative Analysis of Mineral and Energy Resources*, Boston: Reidel, pp. 289–306.

OMRE, H. AND HALVORSEN, K. B. (1989). The Bayesian bridge between simple and universal kriging. *Math. Geology* **21**, 767–786.

OMRE, H., HALVORSEN, K. B. AND BERTEIG, V. (1989). A Bayesian approach to kriging. In Armstrong, M. (ed.), *Geostatistics*, Boston: Kluwer, pp. 109–126.

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

TOLBERT, P., MULHOLLAND, J., MACINTOSH, D., XU, F., DANIELS, D., DEVINE, O., CARLIN, B. P., KLEIN, M., DORLEY, J., BUTLER, A., *et al.* (2000). Air pollution and paediatric emergency room visits for asthma in Atlanta. *American Journal of Epidemiology* **151**, 798–810.

WOODBURY, A. (1989). Bayesian updating revisited. *Math. Geology* **21**, 285–308.

ZHU, L., CARLIN, B. P., ENGLISH, P. AND SCALF, R. (2000). Hierarchical modelling of spatio-temporally misaligned data: relating traffic density to paediatric asthma hospitalizations. *Environmetrics* **11**, 43–61.

[*Received 24 May 2000; revised 12 July 2000 accepted for publication 13 July 2000;*]