Sankhyā : The Indian Journal of Statistics
San Antonio Conference: selected articles
2002, Volume 64, Series A, Pt.2, pp 227-245

PREDICTION, INTERPOLATION AND REGRESSION FOR SPATIALLY MISALIGNED DATA

By S. BANERJEE
University of Minnesota, Minneapolis, USA
and
A.E. GELFAND
University of Connecticut, Storrs, USA

SUMMARY. Spatial models for point-referenced data are used for capturing spatial association and for providing spatial prediction, typically in the presence of explanatory variables. The goal of this paper is to treat the situation where there is misalignment between at least one of the explanatory variables and the response variable. In this context we formalize three inference problems. One, which we call interpolation, seeks to infer about missing response at an observed explanatory location. The second, which we call prediction, seeks to infer about a response at a location with the explanatory variable unobserved. The last, which we call regression, seeks to investigate the functional relationship between the response and explanatory variable through the conditional mean of the response.

We treat both the case of Gaussian and binary spatial response. We adopt a Bayesian approach, providing full posterior inference for each of the above problems. We illustrate both cases using portions of a study of isopod burrows in the Negev desert in Israel.

1. Introduction

Geostatistical problems typically involve inference about a process over a domain based upon a set of measurements taken at a finite number of sites in the domain. In particular, inference about the level of the process at unobserved sites would be vital in such contexts as map development, risk assessment and image analysis. For a *smooth* process, a standard algorthmic approach to this problem adopts a surface approximation given by a parametric function whose form is postulated in advance, either explicitly (e.g., polynomials) or implicitly (e.g., minimum curvature condition).

Paper received October 2000; revised January 2001.
AMS (2000) subject classification. Primary 62M30, 62M20, 60G15.
Keywords and phrases. Bivariate spatial process, cokriging and kriging, covariance function, Gibbs sampler, Kronecker product.

The parameters are selected so as to satisfy some optimality criterion which may be statistical (e.g., least squares) or deterministic (e.g., exact fit at the observed points).

A statistical formulation which introduces a probabilistic specification for spatial association and variability is that of kriging. Kriging is a method for statistical prediction initially proposed by Matheron (1963). Rather than assuming a parametric function for the surface, it is based on arriving at optimal statistical predictors that are linear combinations of the observed data. Prediction of a spatially distributed variable at many unobserved sites enables one to produce a map of this variable. It is clear that, in this formulation, the process need not be smooth. In fact, the variable can be categorical. In the dichotomous case, a binary or two-colour map would result. The literature on classical kriging is very rich, including ordinary kriging, universal kriging, robust kriging, etc. A detailed account is provided in Cressie (1993). By now, there is also a substantial literature on kriging from a partially or fully Bayesian perspective. The work of Omre (1987, 1988), Omre and Halvorsen (1989), Woodbury (1989), Abrahamsen (1993), Hjort and Omre (1994) and Cui, Stein and Myers (1995) is pseudo or empirical Bayes. The more recent work of Handcock and Stein (1993), Handcock and Wallis (1994), De Oliveira, Kedem and Short (1997) and Ecker and Gelfand (1997) is more fully Bayesian.

Often one broadens the investigation of the spatial process of interest to utilize other information available at the observed sites. If the other variables at these sites are viewed as fixed then we can extend the spatial process to include them in a regression fashion. Resulting spatial prediction will reflect this, as in universal kriging (Cressie, 1993, p 151). However, in certain applications, some of the explanatory variables may be viewed as random (stochastic regressors) and modelled most appropriately in a spatial manner. Then, these variables along with the response would result in a multivariate spatial process. Spatial prediction in this framework is usually referred to as cokriging (e.g., Myers, 1982, 1991).

The context of concern to us here involves *misalignment*. The set of observed locations for an explanatory variable is not identical to that for the response variable. Letting Y denote the response and X the explanatory variable, we envision three disjoint subsets of locations in D, the region of interest, R_{XY}, R_{Y} and R_{X}. For locations $\mathrm{s} \in R_{\mathrm{XY}}$ we observe both Y(s) and X(s), for locations in R_{Y} we observe Y(s) but not X(s), for locations in R_{X} we observe X(s) but not Y(s). The remaining locations in D are referred to as ungauged and are labelled as R_{U}. Alternative terminology might refer to our context as one of *missing data*. The $\mathrm{X(s)}'$s are missing on R_{Y}, the $\mathrm{Y(s)}'$s

are missing on R_X. In this regard, such missingness is assumed completely at random. With regard to inference, in order to use all of the data at all of the observed locations, it is clear that X(s) must be viewed as random. We propose to adopt a bivariate spatial process model, in the spirit of the preceding paragraph.

In this context we can formalize three types of inference questions. One concerns Y(s) when $s \in R_X$ which we call interpolation. The second concerns Y(s) for s belonging to R_U, which we call prediction. Evidently, prediction and interpolation are similar but interval estimates will be at least as tight for the latter compared with the former. The last concerns the functional relationship between X(s) and Y(s) at an arbitrary site s, along with other covariate information at s, say $\mathbf{U}(s)$.

In the usual stochastic regressors setting one is interested in the relationship between Y(s) and X(s) where the pairs $(X(s_i), Y(s_i))$, $i = 1, 2, ...,n$ (suppressing $\mathbf{U}(s_i)$), are independent. For us, they are dependent with the dependence captured through a spatial characterization. Still, one may be interested in the regression of Y(s) on X(s) at an arbitrary s. Note that there is no conditional spatial process, $Y(s)|X(s)$, associated with the bivariate spatial process $(X(s), Y(s))$, e.g. how would one define the joint distribution of $Y(s_i)|X(s_i)$ and $Y(s_i')|X(s_i')$?

We also note that our modelling structure differs considerably from that of Diggle, Tawn and Moyeed (1998). They specify a univariate spatial process in order to introduce unobserved spatial effects say V(s) into the modelling after which the $Y(s)'$s are conditionally independent given the $V(s)'$s. In other words, the $V(s)'$s are intended to capture spatial association in the means of the $Y(s)'$s. For us, the $X(s)'$s are also modelled through a spatial process but they are observed and introduced as an explanatory variable with a regression coefficient. Hence, along with the $Y(s)'$s, we require a bivariate spatial process.

Our contribution here is to provide a fully Bayesian examination of the foregoing questions. Initially we study the case where Y(s) is Gaussian but subsequently we allow the response to be binary. The customary advantages to a Bayesian analysis accrue. We obtain an entire posterior for each unknown. Moreover, we can better account for uncertainty in the model specification than empirical Bayes or likelihood based inference can. In particular, even when one can calculate the likelihood exactly, likelihood based inference assumes asymptotics which are generally inappropriate for spatial data (Cressie, 1993, p.350). We can adopt vague priors to let the data drive the inference. Finally we can choose point estimators which retain certain decision theoretic optimalities, e.g., posterior means or medians.

The Gaussian interpolation problem is addressed from an empirical Bayes perspective in a series of papers by Zidek and co-workers. For instance, Le and Zidek (1992) and Brown, Le and Zidek (1994) develop a Bayesian interpolation theory (both spatial and temporal) for multivariate random spatial data. Le, Sun and Zidek (1997) extend this methodology to account for misalignment, i.e. where possibly not all monitored sites measured the same set of pollutants (data missing by design). Their method produces the joint predictive distribution for several locations and different time points using all available data, thus allowing for simultaneous temporal and spatial interpolation without assuming the random field to be stationary. Their approach provides a first stage multivariate normal distribution for the observed data. However, this distribution does not arise from a spatial Gaussian process. Spatial association between measurements at any two different locations is not an explicit function of the pair of spatial locations.

Framing cokriging in the context of linear regression dates to, e.g., Corsten (1989) and Stein and Corsten (1991). In their work, the objective is to carry out predictions for a possible future observation. Stein and Corsten advocate looking at the prediction problem under a regression setup. They propose trend surface modelling of the point source response using polynomials in the coordinates. Typically in trend surface analysis (Cressie, 1993), spatial structure is modelled through the mean but observations are assumed to be independent. Instead, Stein and Corsten (1991) retain familiar spatial dependence structure but assume the resultant covariances and cross-covariances (and hence the dispersion matrix) are known. In this context, Stein, et al. (1991) use restricted maximum likelihood to estimate unknown spatial dependence structure parameters.

Hence the format of the paper is as follows. In Section 2 we provide the development to examine the questions of interest in the Gaussian case. Section 3 provides an enrichment of the covariance structure for a multivariate process model, which may be useful in some applications. Section 4 presents the details for the binary response case. In Section 5 we present illustrative examples for both the Gaussian and binary response cases, using data from a study of isopod burrows in Israeli desert land. Finally, Section 6 presents a summary and possible extensions.

2. The Regression Setup in the Gaussian Case

Geostatistical spatial data customarily refers to measurements on several attributes at point referenced spatial locations, say s_1,s_2, ...,s_n, in a region D.

In the univariate case we observe $Y(s_i)$ at site s_i arising from an underlying random spatial process, $\{Y(s) : s \in D\}$, where D is a fixed subset of R^d with positive measure. That is, the spatial index s varies continuously throughout the region D and a realization of the process is a random surface above D. The data $\{Y(s_1),...Y(s_n)\}$ provides observations of this realization at the locations $\{s_1, ...,s_n\}$. The intrinsic hypothesis (Matheron, 1963) characterizes the process by its mean function $\mu(s) = E(Y(s))$ and its variogram, $2\gamma(h) = var(Y(s+h) - \mu(s+h) - (Y(s) - \mu(s)))$. A stationary Gaussian specification with mean function $\mu(s)$ and covariance function $C(h) = cov(Y(s+h), Y(s))$ implies that the joint distribution of $\mathbf{Y} = (Y(s_1),...Y(s_n))^T$ is multivariate normal. If the covariance function depends only on $\|h\|$, the Euclidean distance between s and s+h, then the processs is called isotropic.

For the regression set up, assuming for the moment a single covariate and no misalignment, let $\mathbf{X} = (X(s_1), ..., X(s_n))^T$ and $\mathbf{Y} = (Y(s_1), ..., Y(s_n))^T$ be the measurements on the covariates and the response, respectively. Assuming it is sensible to model X(s) in a spatial fashion, our approach is to envision a bivariate Gaussian spatial process,

$$\mathbf{W}(s) = \begin{pmatrix} X(s) \\ Y(s) \end{pmatrix} \sim N(\mu(s), T). \tag{1}$$

and assume that the data is realized from this process. Then, the joint distribution of $(\mathbf{X}, \mathbf{Y})^T$ is a multivariate Gaussian distribution. To specify the process in (1) requires modelling the cross-covariances. Our version for the bivariate case immediately extends to the multivariate case. Define the cross-covariance matrix,

$$C(s_1, s_2) = \begin{pmatrix} cov(X(s_1), X(s_2)) & cov(X(s_1), Y(s_2)) \\ cov(Y(s_1), X(s_2)) & cov(Y(s_1), Y(s_2)) \end{pmatrix}. \tag{2}$$

A stationary process results if $C(s_1, s_2)$ depends only upon the separation vector, $s_1 - s_2$.

Modelling for the matrix (2) is a bit more difficult than that associated with a univariate spatial process. The latter only requires that C(h) be a valid covariance function. See e.g. Cressie (1993), Ecker and Gelfand (1997). Now, for locations $s_1, s_2, ..., s_n$, we require the 2nx2n covariance matrix associated with $\mathbf{W} = (\mathbf{W}(s_1), ..., \mathbf{W}(s_n))$ to be positive definite. A convenient constructive approach is to define

$$(C(h))_{lm} = \rho(h;\phi).\tau_{lm}, l = 1, 2; m = 1, 2, \tag{3}$$

where ρ is a valid parametric scalar correlation function and $\mathrm{T} = \begin{pmatrix} \tau_{11} & \tau_{12} \\ \tau_{21} & \tau_{22} \end{pmatrix}$ is positive definite, symmetric. If we collect the set of $\rho(\mathrm{h_{ij}};\ \phi)$, $\mathrm{h_{ij}} = \mathrm{s_i} - \mathrm{s_j}$ into an $n \times n$ matrix H, then we get the covariance matrix for $\mathbf{W}$, $\sum_{\mathbf{W}} = \mathrm{H} \otimes \mathrm{T}$ where $\otimes$ denotes the Kronecker product. The positive definiteness of $\sum_{\mathbf{W}}$ is straightforward. In fact, if $\mathbf{W}(\mathrm{s})$ is $k \times 1$ and T is extended to a $k \times k$ positive definite matrix, $\sum_{\mathbf{W}}$ resulting from (3) is positive definite. This factorized or separable covariance form is considered in Mardia and Goodall (1993) solely in the context of kriging. It is adopted as a prior for $\sum_{\mathbf{W}}$ in Brown, Le and Zidek (1994) and estimated to achieve empirical Bayes kriging. Since $|\sum_{\mathbf{W}}| = |\mathrm{H}|^2 |\mathrm{T}|^{\mathrm{n}}$ and $\sum_{\mathbf{W}}^{-1} = \mathrm{H}^{-1} \otimes \mathrm{T}^{-1}$, likelihood evaluation can be achieved using a 2×2 and an $n \times n$ matrix rather than a $2n \times 2n$ matrix.

In fact, working in the fully Bayesian setting, additional advantages accrue to (3). With ϕ and T a priori independent and an inverse Wishart prior for T, the full conditional distribution for T, that is, $\mathrm{f}(\mathrm{T} \mid \mathbf{W},\ \phi)$, is again an inverse Wishart (e.g., Banerjee, Gelfand and Polasek (2000)). If the Bayesian model is to be fitted using a Gibbs sampler, updating T requires a draw of a 2×2 matrix from a Wishart distribution, substantially faster than updating the $2n \times 2n$ matrix $\sum_{\mathbf{W}}$.

The specification in (3) imposes a possibly inappropriate symmetry in the cross-covariance matrix. This issue and a remedy with considerably increased computational demand are discussed in the next section. Also, (3) presumes that, regardless of the choice of pair of sites, the resulting X covariance is proportional to the resulting Y covariance, i.e., a common ρ and hence a common degree of smoothness for both processes. Evidently, process smoothness cannot be seen from the data though it can be captured using a Matern class for $\rho(\mathrm{h};\ \phi)$. (See, e.g., Stein, (1999).) Hence some preliminary marginal examination of the $\mathrm{X(s)}'\mathrm{s}$ and $\mathrm{Y(s)}'\mathrm{s}$ might help to clarify the feasibility of a common ρ.

Additionally, (3) implies that, for both X's and Y's, correlation between measurements tends to 1 as distance between measurements tends to 0. For some variables, including those in our illustration, such an assumption is appropriate. For others, it may not, whence micro-scale variability, captured through a nugget, is a possible solution. Formally, if independent $\epsilon(\mathrm{s}) \sim N\left(\mathbf{0}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right)$ are added to the modelling of $\mathbf{W}(\mathrm{s})$ in (2), i.e., we write $\mathbf{W}(\mathrm{s}) = \mathbf{V}(\mathrm{s}) + \epsilon(\mathrm{s})$ where $\mathbf{V}(\mathrm{s})$ is distributed as in (2) then the resultant covariance matrix $\sum_{\mathbf{W}} = (\mathrm{H} \otimes \mathrm{T}) + (\mathrm{I} \otimes \mathrm{D})$ where I is $n \times n$ and D

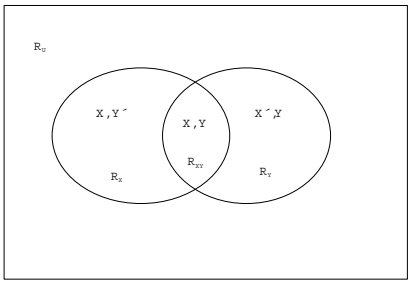

Figure 1. A graphical representation of the R sets. Interpolation applies to locations in R_{X}. Prediction applies to locations in R_{U}. Regression applies to all locations. $\mathbf{X}_{aug} = (\mathbf{X}, \mathbf{X}')$, $\mathbf{Y}_{aug} = (\mathbf{Y}, \mathbf{Y}')$.

is the covariance matrix of $\epsilon(\mathrm{s})$. An increased computational burden results since the full conditional distribution for T is no longer an Inverse Wishart and likelihood evaluation requires working with a $2n \times 2n$ matrix.

With misalignment, $\mathbf{X}$ will be the vector of observed X(s)'s at the sites in $R_{\mathrm{XY}} \cup R_{\mathrm{X}}$ while $\mathbf{Y}$ will be the vector of Y(s)'s at the sites in $R_{\mathrm{XY}} \cup R_{\mathrm{Y}}$. If we let $\mathbf{X}'$ denote the vector of missing X observations in R_{Y} and $\mathbf{Y}'$ the vector of missing Y observations in R_{X}, then in the preceding discussion we can replace $\mathbf{X}$ and $\mathbf{Y}$ by the augmented vectors $\mathbf{X}_{aug} = (\mathbf{X}, \mathbf{X}')$ and $\mathrm{Y}_{aug} = (\mathbf{Y}, \mathbf{Y}')$. See Figure 1 for clarification. After permutation to line up the X's and Y's, they can be collected into a vector $\mathbf{W}_{aug}$. In the Bayesian model specification, $\mathbf{X}'$ and $\mathbf{Y}'$ are viewed as latent (unobserved) vectors. In implementing a Gibbs sampler for model fitting, we update the model parameters given $\mathbf{X}'$ and $\mathbf{Y}'$, i.e., given $\mathbf{W}_{aug}$ and then update $\mathbf{X}'$, $\mathbf{Y}'$ given $\mathbf{X}$, $\mathbf{Y}$ and the model parameters. The latter updating is routine since the associated full conditional distributions are normal. Such augmentation proves computationally easier with regard to bookkeeping since we retain the convenient form for $\sum_{\mathbf{W}}$ below (3). That is, it is easier to marginalize over $\mathbf{X}'$ and $\mathbf{Y}'$ after simulation than before. For convenience of notation, we suppress the augmentation in the sequel.

In what we have called the prediction problem, it is desired to predict the outcome of the response variable at some unobserved site. Thus we are interested in the posterior predictive distribution $[\mathrm{Y}(\mathrm{s}_0) \mid \mathbf{Y}, \mathbf{X}]$. (Here $[\cdot \mid \cdot]$ denotes the conditional density of the variables to the left of $\mid$ given those to the right of $\mid$.) We note that $\mathrm{X}(\mathrm{s}_0)$ is also not observed here. On the other hand, the interpolation problem may be regarded as a method of imputing missing data. Here the covariate $\mathrm{X}(\mathrm{s}_0)$ is observed but the

response is "missing". Thus our attention shifts to the posterior predictive distribution $[Y(s_0) \mid X(s_0), \mathbf{Y}, \mathbf{X}]$. For the regression problem the distribution of interest is $[E(Y(s_0) \mid X(s_0))|X(s_0), \mathbf{Y}, \mathbf{X}]$.

In the absence of any other site level information, we have $\mu(s) = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, independent of the site coordinates. (With additional fixed site level covariates for Y(s) say $\mathbf{U}(s)$, we would replace μ_2 with $\mu_2(s) = \alpha^T \cdot \mathbf{U}(s)$.) Then, from (2), for the pair (X(s),Y(s)), $[Y(s)|X(s), \beta_0, \beta_1, \sigma^2] = N(\beta_0 + \beta_1 X(s), \sigma^2)$, i.e., $E(Y(s)|X(s)) = \beta_0 + \beta_1 X(s)$, where

$$\beta_0 = \mu_2 - \frac{\tau_{12}}{\tau_{11}}\mu_1, \beta_1 = \frac{\tau_{12}}{\tau_{11}}, \sigma^2 = \tau_{22} - \frac{\tau_{12}^2}{\tau_{11}}. \tag{4}$$

So, given samples from the joint posterior distribution of (μ_1, μ_2, T, ϕ), we directly have samples from the posterior distributions for the parameters in (4) and thus from the posterior distribution of $E(Y(s)|X(s))$.

A useful observation in the above modelling framework is that, by rearrangement of the components of **W,**

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \mathbf{1} \\ \mu_2 \mathbf{1} \end{pmatrix}, T \otimes H(\phi)\right) \tag{5}$$

which simplifies calculation of the conditional distribution of $\mathbf{Y} \mid \mathbf{X}$.

Assuming a Wishart prior for T, completing the Bayesian specification requires a prior for μ_1, μ_2, and ϕ. For (μ_1, μ_2), for convenience, we would take a vague but proper bivariate normal prior. A suitable prior for ϕ depends upon the choice of $\rho(h;\phi)$. Then we use a Gibbs sampler to simulate the necessary posterior distributions. The full conditionals for μ_1 and μ_2 are in fact Gaussian distributions, that of the T matrix is inverted Wishart as already mentioned. The full conditional for the ϕ parameter finds ϕ arising in the entries in H and so is not available in closed form. A Metropolis step is employed for its updating.

Under the above framework, interpolation presents no new problems. Let s_0 be a new site at which we would like to predict the variable of interest. We first modify the $H(\phi)$ matrix forming the new matrix H^* as follows,

$$H^*(\phi) = \begin{pmatrix} H(\phi) & \mathbf{h}(\phi) \\ \mathbf{h}(\phi)^T & \rho(0;\phi) \end{pmatrix} \tag{6}$$

where $\mathbf{h}(\phi)$ is the vector with components $\rho(h_{0j}; \phi)$, j = 1, 2, ..., n and h_{0j}

$= s_0 - s_j$. It follows then that

$$\mathbf{W}^* = (\mathbf{W}(s_0), \mathbf{W}(s_1), ..., \mathbf{W}(s_n))^T \sim N\left(\mathbf{1}_{n+1} \otimes \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, H^*(\phi) \otimes T\right). \tag{7}$$

Once again a simple rearrangement of the above vector enables us to arrive at the conditional distribution $[Y(s_0) \mid X(s_0), \mathbf{Y}, \mathbf{X}, \mu, T, \phi]$ as a Gaussian distribution. The predictive distribution for the interpolation problem $[Y(s_0) \mid X(s_0), \mathbf{Y}, \mathbf{X}]$ can now be obtained by marginalizing over the parameters, i.e.,

$$[Y(s_0)|X(s_0), \mathbf{Y}, \mathbf{X}] = \int [Y(s_0)|X(s_0), \mathbf{Y}, \mathbf{X}, \mu, T, \phi][\mu, T, \phi|X(s_0), \mathbf{Y}, \mathbf{X}]. \tag{8}$$

For prediction, we do not have $X(s_0)$. This does not create any new problems as it may be treated as a latent variable and incorporated into $\mathbf{X}'$. This only results in an additional draw within each Gibbs iteration and is a trivial addition to the computational task.

3. Avoiding the Symmetry of the Cross-covariance Matrix

The specification in (3) shows that our model for the cross-covariances imposes a possibly inappropriate symmetry in our cross-covariance matrix. Modelling $\sum_{\mathbf{W}}$ as $H(\phi) \otimes T$ implies that we have

$$\text{cov}(X(s_i), Y(s_j)) = \text{cov}(X(s_j), Y(s_i)) \text{forall} i, j. \tag{9}$$

In the spirit of Le and Zidek (1992) we can avoid this extra symmetry. Instead of directly modelling $\sum_{\mathbf{W}}$ as $H(\phi) \otimes T$ we add a further hierarchical level with $[\sum_{\mathbf{W}}|\phi,T]$ following an inverted Wishart distribution with mean $H(\phi) \otimes T$. All the other specifications remain as before. Note that the marginal model, marginalizing over $\sum_{\mathbf{W}}$, is no longer Gaussian. However, using standard calculations, the cross covariance matrix in (2) is a function of $\rho(h;\phi)$, retaining desirable spatial interpretation. Once again we resort to the Gibbs sampler to arrive at the posteriors, although in this extended model the number of parameters has increased substantially, the elements of $\sum_{\mathbf{W}}$ being introduced as new parameters.

The full conditionals for the means μ_1 and μ_2 are still Gaussian and it is easily seen that the full conditional for $\sum_{\mathbf{W}}$ is inverted Wishart. The full conditional distribution for ϕ is now proportional to $[\sum_{\mathbf{W}}|\phi,T] \cdot [\phi]$; a

Metropolis step is employed for its updating. Also, the full conditional for T is no longer inverted Wishart and a Metropolis step with an inverted Wishart proposal is used to sample the T matrix. All told, this is indeed a much more computationally demanding proposition since we now have to deal with the $2n \times 2n$ matrix $\sum_{\mathbf{W}}$ with regard to sampling, inversion, determinant, etc. Nonetheless, this modelling was implemented for the first example in Section 5. In this example, inference was essentially indistinguishable from that of Section 2 and so is not reported.

4. Regression Setup in a Probit model

Now suppose we have binary response from a point-source spatial dataset. At each site, Z(s) equals 0 or 1 according to whether we observed "failure" or "success" at that particular site. Thus, a realization of the process can be partitioned into two disjoint subregions, one for which Z(s) = 0, the other Z(s) = 1, and is called a binary map (De Oliveira, 2000). Again, the process is only observed at a finite number of locations. Along with this binary response we have a set of covariates observed at each site. We follow the latent variable approach for probit modelling as in, e.g., De Oliveira (2000). Let Y(s) be a latent spatial process associated with the sites and let X(s) be a process that generates the values of a particular covariate, in particular, one that is misaligned with Z(s) and is sensible to model in a spatial fashion. For the present we assume X(s) is univariate but extension to the multivariate case is apparent. Let Z(s) = 1 i.f.f. Y(s) > 0. We envision our bivariate process as $\mathbf{W}(\mathrm{s}) = \begin{pmatrix} \mathrm{X}(\mathrm{s}) \\ \mathrm{Y}(\mathrm{s}) \end{pmatrix} \sim \mathrm{N}(\mu(\mathrm{s}), \mathrm{T})$, as in (1) where now $\mu(\mathrm{s}) = \begin{pmatrix} \mu_1 \\ \mu_2 + \alpha^T \mathbf{U}(\mathrm{s}) \end{pmatrix}$ with $\mathbf{U}(\mathrm{s})$ regarded as a $p \times 1$ vector of fixed covariates. Note that the conditional variance of Y(s) given X(s) is not identifiable. Thus, without loss of generality we set $\tau_{22} = 1$ so that the T matrix has only two parameters.

Now, we formulate a probit regression model as follows.

$$P(\mathrm{Z}(\mathrm{s}) = 1 | \mathrm{X}(\mathrm{s}), \mathbf{U}(\mathrm{s}), \alpha, \mu_1, \mu_2, \tau_{11}, \tau_{12}) = \Phi\left(\frac{\beta_0 + \beta_1 \mathrm{X}(\mathrm{s}) + \alpha^T \mathbf{U}(\mathrm{s})}{\left(1 - \frac{\tau_{12}^2}{\tau_{11}}\right)^{1/2}}\right) \tag{10}$$

Here, as in (4), $\beta_0 = \mu_2 - \frac{\tau_{12}}{\tau_{11}}\mu_1$, $\beta_1 = \frac{\tau_{12}}{\tau_{11}}$.

The posterior of interest is $[\mu_1, \mu_2, \alpha, \tau_{11}, \tau_{12}, \phi, \mathbf{Y} \,|\mathbf{X}, \mathbf{Z}]$ where $\mathbf{Z} = (\mathrm{Z}(\mathrm{s}_1), ..., \mathrm{Z}(\mathrm{s}_\mathrm{n}))^T$ is a vector of 0's and 1's. The fitting again uses Markov chain Monte Carlo. Here, $\mathbf{X} = (\mathrm{X}(\mathrm{s}_1), ..., \mathrm{X}(\mathrm{s}_\mathrm{n}))^T$ and $\mathbf{Y} = (\mathrm{Y}(\mathrm{s}_1), ..., \mathrm{Y}(\mathrm{s}_\mathrm{n}))^T$ as in Section 2, except that $\mathbf{Y}$ is now unobserved, introduced for computational convenience. Analogous to (5),

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_1 \mathbf{1} \\ \mu_2 \mathbf{1} + \mathbf{U}\beta \end{pmatrix}, \mathrm{T} \otimes \mathrm{H}(\phi) \right) \tag{11}$$

where $\mathbf{U} = (\mathbf{U}(\mathrm{s}_1), ..., \mathbf{U}(\mathrm{s}_\mathrm{n}))^T$.

From (11), the full conditional distribution for each latent $\mathrm{Y}(\mathrm{s}_\mathrm{i})$ is a univariate normal truncated to a set of the form $(\mathrm{Y}(\mathrm{s}_\mathrm{i}) > 0)$ or $(\mathrm{Y}(\mathrm{s}_\mathrm{i}) < 0)$. The full conditionals for μ_1 and μ_2 are both univariate normal and that of β is multivariate normal with the appropriate dimension. For the elements of the T matrix we may simulate first from a Wishart distribution (as mentioned in Section 2) and then proceed to scale it by τ_{22} or we may proceed individually for τ_{12} and τ_{11} using Metropolis Hastings over a restricted convex subset of a hypercube (Chib and Greenberg, 1998). ϕ is simulated using a Metropolis step, as in Section 2. Misalignment is treated as in Section 2, introducing appropriate latent $\mathbf{X}'$ and $\mathbf{Y}'$.

With posterior samples from $[\mu_1, \mu_2, \alpha, \tau_{11}, \tau_{12}, \phi|\mathbf{X}, \mathbf{Z}]$, we immediately obtain samples from the posterior distributions for β_0 and β_1. Also, given $\mathrm{X}(\mathrm{s}_0)$, (10) shows how to obtain samples from the posterior for a probability such as $[P(\mathrm{Z}(\mathrm{s}_0) = 1|\mathrm{X}(\mathrm{s}_0), \mathbf{U}(\mathrm{s}_0), \alpha, \mu_1, \mu_2, \tau_{11}, \tau_{12}) \mid \mathrm{X}(\mathrm{s}_0), \mathbf{X}, \mathbf{Z}]$ at unobserved site s_0, providing further clarification to the regression structure. Were $\mathrm{X}(\mathrm{s}_0)$ not observed, we could still consider the chance that $\mathrm{Z}(\mathrm{s}_0) = 1$. This probability arises by averaging over $\mathrm{X}(\mathrm{s}_0)$, i.e.,
$P(\mathrm{Z}(\mathrm{s}_0) = 1|\mathbf{U}(\mathrm{s}_0), \alpha, \mu_1, \mu_2, \tau_{11}, \tau_{12}) =$

$$\int P(\mathrm{Z}(\mathrm{s}_0) = 1|\mathrm{X}(\mathrm{s}_0), \mathbf{U}(\mathrm{s}_0), \alpha, \mu_1, \mu_2, \tau_{11}, \tau_{12}) [\mathrm{X}(\mathrm{s}_0)|\mu_1, \tau_{11}]. \tag{12}$$

In practice, we would replace the integration in (12) by a Monte Carlo integration. Then, plugging into this Monte Carlo integration, the foregoing posterior samples would yield essentially posterior realizations of (12).

Both the prediction problem and the interpolation problem may be viewed as examples of indicator kriging (e.g., Solow, 1986, De Oliveira, 2000). For the prediction case we seek $[\mathrm{Z}(\mathrm{s}_0)|\mathbf{X}, \mathbf{Z}]$. Realizations arise if we can obtain realizations from $[\mathrm{Y}(\mathrm{s}_0)|\mathbf{X}, \mathbf{Z}]$. But

$$[\mathrm{Y}(\mathrm{s}_0)|\mathbf{X}, \mathbf{Z}] = \int [\mathrm{Y}(\mathrm{s}_0)|\mathbf{X}, \mathbf{Y}] [\mathbf{Y}|\mathbf{X}, \mathbf{Z}]. \tag{13}$$

Since the first distribution under the integral in (13) is a univariate normal, as in Section 2, the posterior samples of $\mathbf{Y}$ immediately provide samples of $\mathrm{Y}(\mathrm{s}_0)$. For the interpolation case we seek $[\mathrm{Z}(\mathrm{s}_0)|\mathrm{X}(\mathrm{s}_0), \mathbf{X}, \mathbf{Z}]$. Again we only need realizations from $[\mathrm{Y}(\mathrm{s}_0)|\mathrm{X}(\mathrm{s}_0), \mathbf{X}, \mathbf{Z}]$. But

$$[\mathrm{Y}(\mathrm{s}_0)|\mathrm{X}(\mathrm{s}_0), \mathbf{X}, \mathbf{Z}] = \int [\mathrm{Y}(\mathrm{s}_0)|\mathrm{X}(\mathrm{s}_0), \mathbf{X}, \mathbf{Y}][\mathbf{Y}|\mathrm{X}(\mathrm{s}_0), \mathbf{X}, \mathbf{Z}]. \tag{14}$$

As with (13), the first distribution under the integral in (14) is a univariate normal.

5. Illustrative Examples

Our examples are based upon an ecological dataset collected over a west-facing watershed in the Negev desert in Israel. The species under study is called an isopod and builds its residence by making burrows. Some of these burrows thrive through the span of a generation while others do not. We study the following variables at each of 1129 sites. The variable dew measures time in minutes (from 8 am) to evaporation of dew. The variables shrub and rock density are percentages (the remainder is sand) characterizing the environment around the burrows. In our first example we try to explain shrub density (Y) through dew duration (X). In our second example we try to explain burrow survival (Z) through shrub density, rock density and dew duration treating only the last one as random and spatial. We illustrate the Gaussian case for the first example with 714 of the sites offering both measurements, 204 sites providing only the shrub density and 211 containing only the dew measurements.

The spatial locations are displayed in Figure 2 using rescaled planar coordinates after UTM projection. The rectangle in Figure 2 is roughly 300 kms by 250 kms. Of the 714 sites providing information on both variables, 20 sites were selected at random and were not incorporated into the analysis but rather were used for cross-validation. Hence the vector $\mathbf{X}$ consists of $694 + 211 = 905$ measurements and the vector $\mathbf{Y}$ consists of $694 + 204 = 898$ measurements. Also, for these examples we take the correlation function $\rho(\mathrm{h}; \phi) = \mathrm{e}^{-\phi \mathrm{h}}$ so that $\frac{3}{\phi}$ is interpreted as the range of spatial association (Ecker and Gelfand, 1997). For this correlation function, we assign a vague inverse gamma specification for the parameter ϕ, i.e. IG(2, 0.024). This prior has infinite variance and suggests a range of 125 kms which is roughly half the maximum pairwise distance in our region. We have found little inference sensitivity to the mean of this prior. The remaining prior specifications

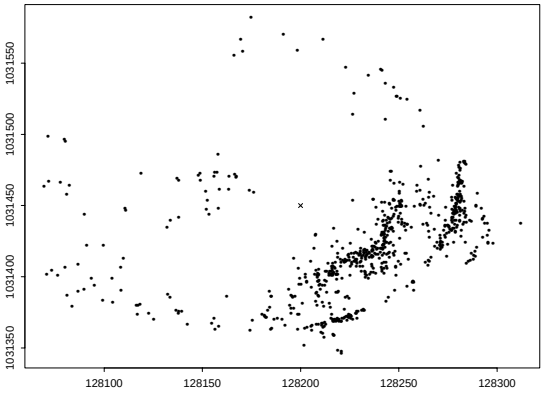

Figure 2. Spatial locations of the isopod burrows data in Section 5. The axes represent the Eastings and the Northings on an UTM projection.

are all rather non-informative, i.e., a $\mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10^5 & 0 \\ 0 & 10^5 \end{pmatrix}\right)$ prior for (μ_1,μ_2) and an $\mathrm{IW}\left(2, \begin{pmatrix} 0.001 & 0 \\ 0 & 0.001 \end{pmatrix}\right)$ for T. That is, $\mathrm{E}(\tau_{11}) = \mathrm{E}(\tau_{22}) = 0.001$, $\mathrm{E}(\tau_{12}) = 0$ and the variances of the τ's do not exist.

Table 1 provides the 95% credible intervals for the regression parameters and the geostatistical range parameter ϕ. The significant negative association between dew duration and shrub density is unexpected but is evident on a scatterplot of the 714 sites having both measurements.The intercept β_0 is significantly high while the slope β_1 is negative. Also the maximum distance in the sample is approximately 248.1 kms so the spatial range, computed from the point estimate of $\frac{3}{\phi}$ from Table 1 is approximately 99.7 kms is about 40% of the maximum distance. In Figure 3(a) we show the relative performances (using posterior density estimates) of prediction, interpolation and regression at a somewhat central location s_0 indicated by a $\times$ in Figure 2. The associated $\mathrm{X}(\mathrm{s}_0)$ has the value 73.10 minutes. Regression, since it models the means rather than predicting a variable, has substantially smaller variability than prediction and interpolation. In Figure 3(b) we compare the latter pair. As is expected, interpolation has less variability due to the specification of $\mathrm{X}(\mathrm{s}_0)$. In Table 2, we present our cross-validation results for the twenty sites that we had removed from our analysis, providing prediction and interpolation intervals for the response, along with the actual observed shrub density measurement at each of the sites. It turns out that, in all cases the observed value falls within the associated intervals. Finally, in Fig-

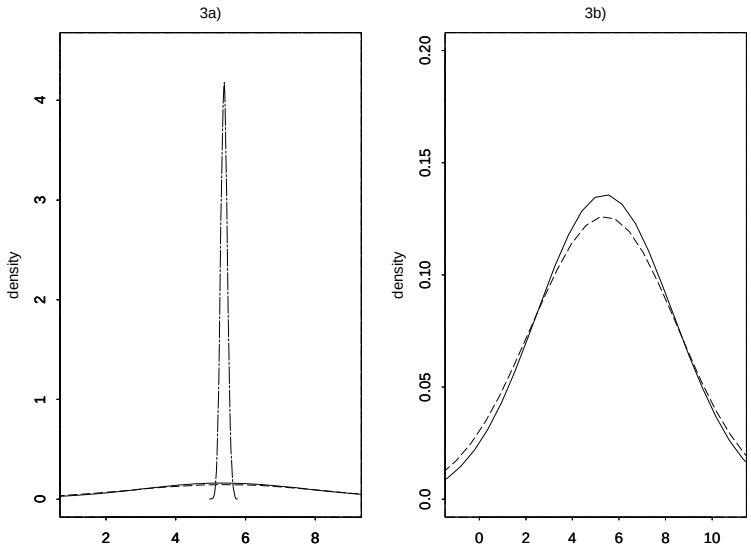

Figure 3. Posterior distributions for inference at the location s_0, denoted by $\times$ in Figure 2.
– - – denotes $[E(Y(s_0)|X(s_0))|\mathbf{X},\mathbf{Y}]$,
—— denotes $[Y(s_0)|(s_0),\mathbf{X},\mathbf{Y}]$ and
– – – denotes $[Y(s_0)|\mathbf{X},\mathbf{Y}]$.

ure 4 we present a three-dimensional surface plot of E(Y(s)|**Y**,**X**) over the region. This plot reveals the spatial pattern in shrub density over the watershed. Higher measurements are expected in the eastern part, particularly the southeastern part, of the region while relatively fewer shrubs are found in the northern and western parts.

Table 1. Inference summary for the shrub density / dew duration example discussed in Section 5. See text for details.

	Quantiles		
Parameter	2.5%	50%	97.5%
μ_1	73.118	73.885	74.665
μ_2	5.203	5.383	5.572
τ_{11}	95.095	105.220	117.689
τ_{12}	-4.459	-2.418	-0.528
τ_{22}	5.564	6.193	6.914
ρ (Corr. Coeff.)	-0.171	-0.095	-0.021
β_0 (Intercept)	5.718	7.078	8.463
β_1 (Slope)	-0.041	-0.023	-0.005
σ^2	5.582	6.215	6.931
ϕ	0.0091	0.0301	0.2072

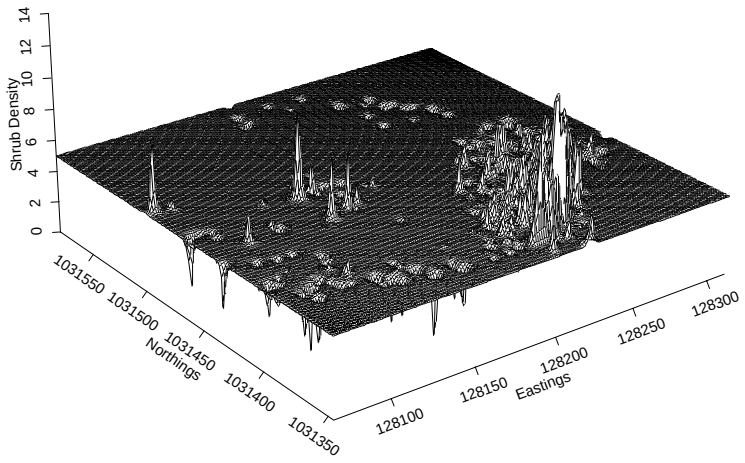

Figure 4. For the Gaussian analysis case, a three dimensional surface plot of $E(Y(s)|\mathbf{X}, \mathbf{Y})$ over the region in Figure 2.

Table 2. Cross-validaton summary for the shrub density / dew duration example discussed in Section 5. See text for details

Site #	Actual Value	Prediction	Interpolation
1	5.36	(0.651, 10.210)	(1.595, 9.277)
2	4.79	(0.640, 10.120)	(0.666, 9.492)
3	8.67	(1.461, 10.127)	(1.906, 10.182)
4	4.15	(0.666, 10.176)	(0.336, 8.828)
5	6.45	(0.497, 10.140)	(0.827, 10.069)
6	7.03	(0.502, 10.196)	(1.014,10.362)
7	3.96	(0.646, 10.124)	(0.604, 10.389)
8	5.07	(0.623, 10.429)	(0.968, 10.452)
9	0.76	(0.418, 10.345)	(0.638, 9.361)
10	4.76	(0.617, 10.205)	(0.457, 8.460)
11	3.18	(0.439, 9.308)	(0.491, 8.061)
12	6.03	(0.625, 10.334)	(0.550, 10.127)
13	2.94	(0.442, 9.246)	(0.375, 7.975)
14	4.29	(0.637, 10.269)	(1.566, 10.271)
15	5.24	(0.657, 10.264)	(1.917, 10.135)
16	7.41	(0.807, 10.271)	(1.721, 10.672)
17	3.93	(0.647, 10.165)	(0.161, 8.262)
18	6.77	(0.674, 10.144)	(1.254, 10.578)
19	4.69	(0.648, 10.532)	(0.750, 10.431)
20	3.55	(0.706, 11.200)	(1.862, 9.858)

Our second example uses a smaller dataset, from the same region as Figure 2, which has 246 burrows of which 43 do not provide the dew measurements. Here the response is binary, governed by the success (Y=1) or failure of a burrow at a particular site. The explanatory variables, dew duration, shrub density and rock density relate, in some fashion, to water retention. Dew measurements are taken as the X's in our modelling with shrub and rock density being U_1 and U_2 respectively. The prior specifications leading to the probit modelling again have vague bivariate normal priors for (μ_1, μ_2) and so also for β which is two dimensional in this example. For ϕ we again assign a non-informative inverse gamma specification, IG(2, 0.024). We generate τ_{11} and τ_{12} through scaling a Wishart distribution for T with prior, $\mathrm{IW}\left(2, \begin{pmatrix} 0.001 & 0 \\ 0 & 0.001 \end{pmatrix}\right)$. In Table 3, we present the 95% credible intervals for the parameters in the model. The positive coefficient for dew is expected. It is interesting to note that shrub and rock density seem to have a negative impact on the success of the burrows. This leads us to believe that though high shrub and rock density may encourage the hydrology it is perhaps not conducive to the growth of food materials for the isopods or encourages predation of the isopods. The spatial range parameter again explains about 40% of the maximum distance. In Figure 5 we present the density estimates for the posteriors $[P(\mathrm{Z}(\mathrm{s}_0) = 1|\mathrm{X}(\mathrm{s}_0), \mathbf{U}(\mathrm{s}_0), \alpha, \beta_0, \beta_1)|\mathrm{X}(\mathrm{s}_0), \mathbf{X}, \mathbf{Z}]$ and $[P(\mathrm{Z}(\mathrm{s}_0) = 1|\mathbf{U}(\mathrm{s}_0), \alpha, \mu_1, \mu_2, \tau_{11}, \tau_{12})|\mathbf{X}, \mathbf{Z}]$ with s_0 being a rather central location and $\mathrm{X}(\mathrm{s}_0) = 74.7$ minutes (after 8:00 am), to compare performance of interpolation and prediction. As expected, interpolation provides a slightly tighter posterior distribution than prediction.

Table 3. Inference summary for burrow survival data example from Section 5. See text for details

	Quantiles		
Parameter	2.5%	50%	97.5%
μ_1	75.415	76.095	76.772
μ_2	0.514	1.486	2.433
τ_{11}	88.915	99.988	108.931
τ_{12}	0.149	0.389	0.659
ϕ	0.0086	0.0302	0.2171
β_0 (Intercept)	0.310	1.256	2.200
β_1 (Dew slope)	0.032	0.089	0.145
α_1 (Shrub)	-0.0059	-0.0036	-0.0012
α_2 (Rock)	-0.00104	-0.00054	-0.00003

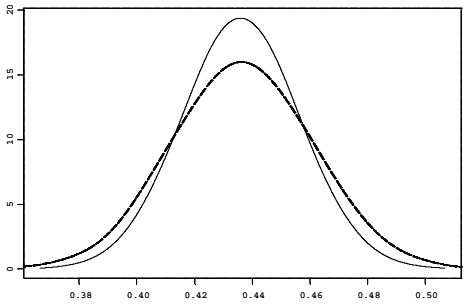

Figure 5. For the Probit data analysis in Section 5, posterior $P(Z(s_0) = 1|X(s_0), \mathbf{U}(s_0), \alpha, \beta_0, \beta_1)$ indicated by —— and posterior $P(Z(s_0) = 1|\mathbf{U}(s_0), \alpha, \mu_1, \mu_2, \tau_{11}, \tau_{12})$ indicated by $- - -$. See text for details.

6. **Summary and Extensions**

Often point-referenced spatial data layers are misaligned. That is, with two layers, calling one, say Y, the response variable, the other, say X, the explanatory variable, there are some locations where X is observed but not Y and vice versa. In this context three inference problems may be of interest. One is the regression relationship at location s, captured by $E(Y(s)|X(s))$. The other two are interpolation - inference about Y(s) at a location s where X(s) is observed - and prediction - inference about Y(s) where X(s) is unobserved. We have provided a fully Bayesian approach to these problems for the cases where Y(s) is from a Gaussian process and a binary process. Our approach is based upon a bivariate stationary Gaussian process model., assuming a separable cross-covariance function.

Extension to accommodate additional misaligned covariates is evident. A computationally manageable approach to accommodate nonstationarity can be implemented by extending the recent locally-stationary modeling strategy of Fuentes (2000). Finally, handling the spatio-temporal setting can be accomplished in two distinct ways. One extends (3) to introduce time in a separable fashion following Mardia and Goodall (1993). The other discretizes time and then nests space within time introducing a spatial process for each time value as in Gelfand, et al. (2000).

References

Abrahamsen, P. (1993). Bayesian Kriging for Seismic Depth Conversion of a Multi Layer Resorvoir. In *Geostatistics Troia*, A. Soares, ed., 1992. Kluwer Academic Publishers. 385-398.

Banerjee, S., Gelfand, A.E. and Polasek, W. (2000). Geostatistical Modeling for Spatial Interaction Data with Application to Postal Service Performance. *J. Statist. Plann. Inference*, **90**, 87-105.

Brown, P.J., Le, N.D. and Zidek, J.V. (1994). Multivariate Spatial Interpolation and Exposure to Air Pollutants. *Canadian J. Statist.*, **22**, 489-510.

Chib, S. and Greenberg, E. (1998). Analysis of Multivariate Probit Models. *Biometrika*, **85**, 347-361.

Corsten, L.C.A. (1989). Interpolation and optimal linear prediction. *Statist. Neerlandica.* **43**, 69-84. Corrigenda (1990), **44**, i.

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

Cui, H., Stein, A. and Myers D. (1995). Extension of Spatial Information, Bayesian Kriging and Updating of Prior Variogram Parameters. *Environmetrics.* **6,** 373-384.

De Oliveira, V. (2000). Bayesian prediction of Clipped Gaussian Random Fields, Computational Statistics and Data Analysis. (to appear)

De Oliveira, V., Kedem B. and Short, D. (1997). Bayesian Prediction of transformed Gaussian Random Fields. *J. Amer. Statist. Assoc.*, **92**, 1422-1433.

Diggle, P.J., Tawn, J.A. and Moyeed, R.A. (1998). Model Based Geostatistics (with discussion), *Appl. Statist.*, **47**, 299-350.

Ecker, M.D. and Gelfand, A.E. (1997). Bayesian variogram modeling for an isotropic spatial process. *J. Agric. Biol. Environ. Stat.*, **2**, 347-369.

Fuentes, M. (2000). A New High Frequency Kriging Approach for Nonstationary Environmental Processes. Technical Report, Department of Statistics, North Carolina State University.

Gelfand, A.E., Ecker, M.D., Knight, J.R. and Sirmans, C.F. (2000). The Dynamics of location in home price. Technical Report, Department of Statistics, University of Connecticut.

Handcock, M. and Stein, M. (1994). A Bayesian Analysis of Kriging. *Technometrics,* **35** (4), 403-410.

Handcock, M. and Wallis, J. (1994). An approach to Statistical Spatial-Temporal modelling of Metereological Fields (with Discussion).*J. Amer. Statist. Assoc.,* **89**, 368-390.

Hjort, N. and Omre, H. (1994). Topics in Spatial Statistics (with Discussion). *Scand. J. Statist.*, **21**, 289-357 .

Le, N.D. and Zidek, J.V. (1992). Interpolation with uncertain spatial covariance: a Bayesian alternative to kriging. *J. Multivariate Anal.*, **43,** 351-374.

Le, N.D., Sun, W. and Zidek, J.V. (1997). Bayesian multivariate spatial interpolation with data missing by design. *J. Roy. Statist. Soc. Ser. B,* **59,** 501-510.

Mardia, K.V. and Goodall, C.R. (1993). Spatial-temporal analysis of multivariate environmental monitoring data. In: *Multivariate Environmental Statistics.* G.P. Patil and C.R. Rao, eds., Elsevier, Amsterdam, 347-386.

Matheron, G. (1963). Principles of Geostatistics. *Economic Geology*, **58**, 1246-1266.

Myers, D.E. (1982). Matrix formulation of cokriging, *J. International Assoc. Math. Geology,* **14**, 249-257.

MYERS, D.E. (1991). Pseudo cross-variograms, positive-definiteness and cokriging. *Math. Geology.* **23**, 805-816.

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 *Quantitative Analysis of Mineral and Energy Resources.* C. F. Chung et al., eds., D. Reidel Publishing Co., Boston, 289-306.

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

SOLOW, A.R. (1986). Mapping by simple indicator kriging. *Math. Geology*, **18**, 335-354.

STEIN, A. and CORSTEN, L.C.A. (1991). Universal kriging and cokriging as a regression procedure. *Biometrics,* **47**, 575-587.

STEIN, A., VAN EIJNBERGEN, A.C. and BARENDREGT, L.G. (1991). Cokriging Nonstationary Data. *Math. Geology.* **23**, 703-719.

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

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

S. BANERJEE
DIVISION OF BIOSTATISTICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN
E-mail: sudiptob@biostat.umn.edu

A.E. GELFAND
DEPARTMENT OF STATISTICS
UNIVERSITY OF CONNECTICUT
STORRS, CT
E-mail: alan@stat.uconn.edu