Spatial prediction and temporal backcasting for environmental fields having monotone data patterns

Nhu D. LE, Li SUN and James V. ZIDEK

Abstract: The authors develop a methodology for predicting unobserved values in a conditionally lognormal random spatial field like those commonly encountered in environmental risk analysis. These unobserved values are of two types. The first come from spatial locations where the field has never been monitored, the second, from currently monitored sites which have been only recently installed. Thus the monitoring data exhibit a monotone pattern, resembling a staircase whose highest step comes from the oldest monitoring sites. The authors propose a hierarchical Bayesian approach using the lognormal sampling distribution, in conjunction with a conjugate generalized Wishart distribution. This prior distribution allows different degrees of freedom to be fitted for individual steps, taking into account the differential amounts of information available from sites at the different steps in the staircase. The resulting hierarchical model is a predictive distribution for the unobserved values of the field. The method is demonstrated by application to the ambient ozone field for the southwestern region of British Columbia.

Prévisions spatiales et temporelles pour des champs de données environnementales disposées en escalier

Résumé : Les auteurs développent une méthode permettant de faire de la prévision pour des valeurs non-observées dans un champ spatial aléatoire conditionnellement lognormal comme ceux que l'on rencontre fréquemment dans l'analyse des risques environnementaux. Les valeurs non-observées correspondent soit à des sites où le champ n'a jamais été mesuré, soit à des stations d'observation en activité mais qui n'ont été installées que récemment. On peut donc distinguer dans les données un patron monotone semblable à un escalier dont la plus haute marche correspond aux sites les plus anciens. Les auteurs proposent une approche bayésienne hiérarchique s'appuyant sur la loi lognormale et une loi a priori de Wishart généralisée conjuguée. Cette dernière autorise l'emploi de degrés de liberté différents pour chacune des marches, ce qui permet de refléter la quantité d'information disponible à chacun des sites le long de l'escalier. On peut alors déduire de ce modèle hiérarchique une loi prévisionnelle pour les valeurs non-observées du champ. L'approche est illustrée à l'aide de données sur le niveau ambiant d'ozone dans le sud-ouest de la Colombie-Britannique.

1. INTRODUCTION

Risk assessments of air pollution often require estimates of the concentration levels at locations where there are no monitoring sites. Such assessments may rely on data from monitored locations. Spatial prediction methods should then be used to accomplish what Carroll, Ruppert & Stefanski (1995) call "regression calibration," a technique which Pierce, Stram, Vaeth & Schafer (1992) advocate as a way to reduce the potentially deleterious effects of measurement error. Such methods predict unmeasured air pollution levels at locations where people live, based on observed concentration levels at monitored sites. Such an approach has been used elsewhere (see Duddek, Le, Zidek & Burnett 1995; Zidek 1997; Zidek, White, Le, Sun & Burnett 1998).

For chronic diseases such as cancer with long latency periods where cumulative exposure seems relevant, predicted concentrations are needed for long periods of time. Hence observed concentrations at some time periods may need to be used for predicting those at others. For example, in a case-control study, the residential histories of the participants are determined so

that their cumulative exposure estimates may be found by predicting concentrations at their former residential locations. As well, in this case, levels may be predicted at currently monitored locations for time periods before their monitors were installed. We call making such predictions "backcasting".

In recent years, a Bayesian methodology for both temporal and spatial prediction of air pollution has been developed, beginning with the work of Le & Zidek (1992). That method was seen as an alternative to kriging (see Cressie 1991). Its hierarchical prior model seemed to offer more flexibility than kriging, originally developed for interpolating fixed fields in geostatistics rather than the stochastic fields encountered in atmospheric science. Moreover, it was seen to offer a way of incorporating the uncertainty inevitably associated with the components of complex models.

Ensuing development confirmed the early promise. The method can be enhanced to deal with the realistic multivariate setting where not all monitoring sites measure the same set of pollutants (Brown, Le & Zidek 1994a; Le, Sun & Zidek 1997). The method produces the joint predictive distribution for several locations and different time points using all available data, thus allowing the simultaneous temporal and spatial prediction. Moreover, unlike kriging, the method does not require the random field to be spatially stationary. Furthermore, it allows for the uncertainty associated with the mean and the spatial covariance of the field to be incorporated in the predictive distribution.

Empirical comparisons (see Sun 1998) validate the method. In particular, the analysis of Sun (1998) indicates that the Bayesian predictor will outperform kriging. At least it outperforms a cokriging variant of kriging suggested by Haas (1996) in a multivariate setting. The Bayesian alternative produces predictive intervals with almost correct coverage probabilities while the cokriging-based prediction intervals yield coverage probabilities much lower than nominal.

One important restriction in these methodological developments is that all monitoring stations are required to be in operation for the same period. This restriction limits the flexibility of the methodology in that stations might be added over time to monitoring networks; data collected from networks with different operational periods cannot be integrated into a single analysis.

In this paper, we address the practically important problem of temporal-spatial interpolation where stations have been added over time to a monitoring network. After appropriately reordering the sites, the resulting data matrix will then have a staircase structure. That is, when the data are put together in an increasing order of the stations' operational periods, the data matrix has the appearance of a staircase. Combining active networks with different starting times will yield such a structure.

The staircase pattern can be seen in the air pollution data from monitoring networks maintained by the Ministry of Environment in the province of British Columbia (BC), Canada. Figure 1 shows the locations along with their starting times and boxplots of the monthly average ozone levels. Here, each step of the staircase consists of stations having the same starting time. These data are the subject of the application considered in the next section.

The staircase structure is also described as a monotone missing data pattern by some authors (Little & Rubin 1987; Rubin & Shaffer 1990; Liu 1999) who have studied it in the context of data imputation. In particular, an approach to obtaining parameter estimates in a Bayesian multivariate linear regression problem with such missing pattern has been studied by Liu (1996).

In this paper, we develop a Bayesian method for temporal-spatial interpolation using all available data with this special feature. The method relies on the generalized inverted Wishart distribution (see Brown, Le & Zidek 1994b). Specifically, the monthly averages of ozone are assumed to follow a Gaussian distribution and the corresponding covariance is assumed to follow a generalized inverted Wishart (GIW) prior distribution.
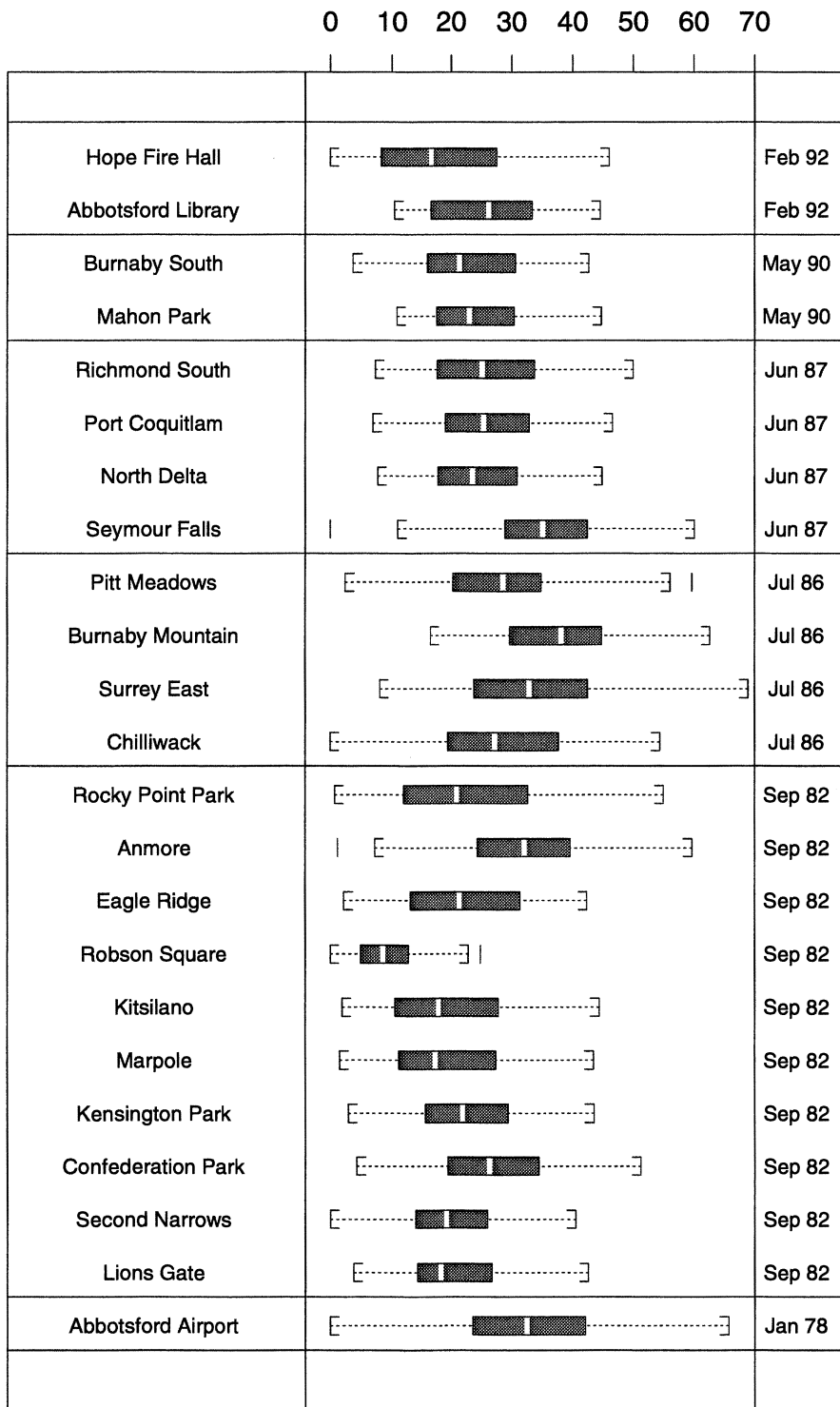
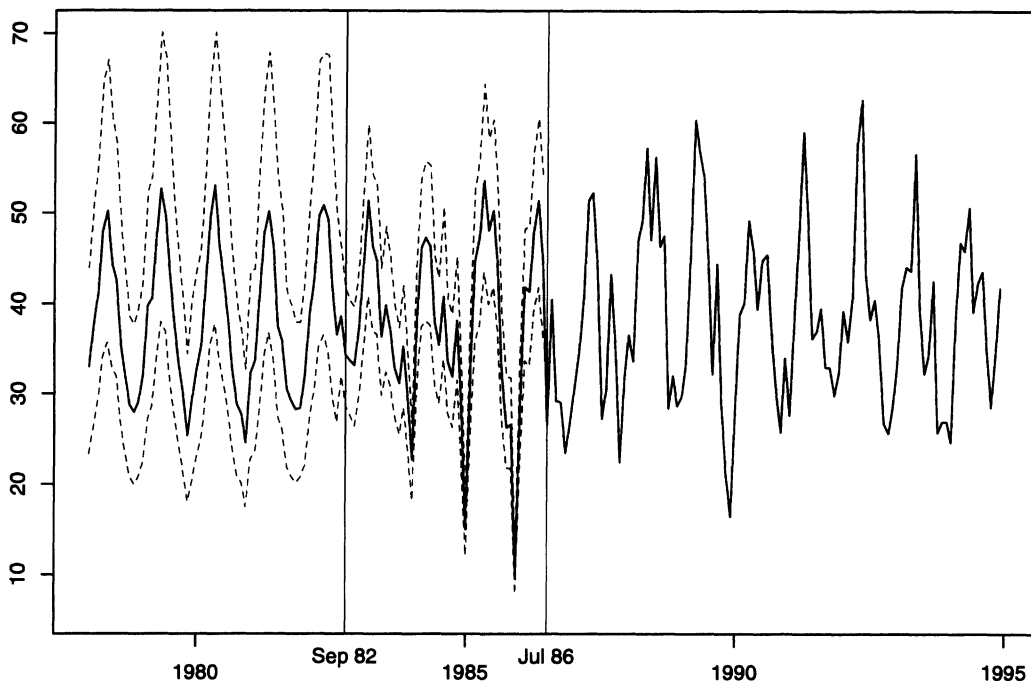FIGURE 1: Boxplot of monthly average ozone levels ($\mu g/m^3$) at 23 monitor sites in British Columbia, Canada, and their start-up times.

FIGURE 2: Observed monthly average ozone levels in $\mu g/m^3$ at the Burnaby Mountain station between July 1986 and December 1994. Backcast values (solid) and corresponding 95% predictive intervals (dash) between January 1978 and June 1986. The vertical lines indicate when blocks are formed during this backcasting period.

The methodology we develop yields the joint predictive distribution for several locations and time points using all the observed data in the staircase. The method extends the Bayesian methods discussed above and thus enjoys all the corresponding advantages. The method is applied to the BC data cited above. From the resulting predictive distribution, it is possible to get the temporal predictions and the corresponding predictive intervals for any locations of interest in the spatial field, e.g., as displayed in Figure 2. The method is seen to work reasonably well, in that the predicted means capture not only the temporal features demonstrated at specific stations but also spatial features of the pollution field.

The paper is organized as follows. We demonstrate the use of our new method in Section 2. The main theoretical results are described in Section 3 with relevant proofs given in the appendices. Parameter estimation is discussed in Section 4. Concluding remarks then follow in Section 5.

2. APPLICATION

2.1. Data set.

We illustrate the methodology through an application to ozone data. The data set consists of monthly average ozone levels over 23 monitor sites in the southwestern region of Canada's westernmost province, British Columbia (BC). The monthly average levels, provided to us by the BC Ministry of Environment, were calculated as follows. First, daily averages were computed by averaging over hourly measurements for days with at least 18 hours of valid measurements. The monthly averages were then calculated based on these daily averages for those months with at most 7 days of missing values from January 1978 to December 1994. Many sites started operation later than 1978, and the stations having the same starting time are grouped together to create steps in the staircase in Figure 1. The locations of these sites are shown in Figure 3.
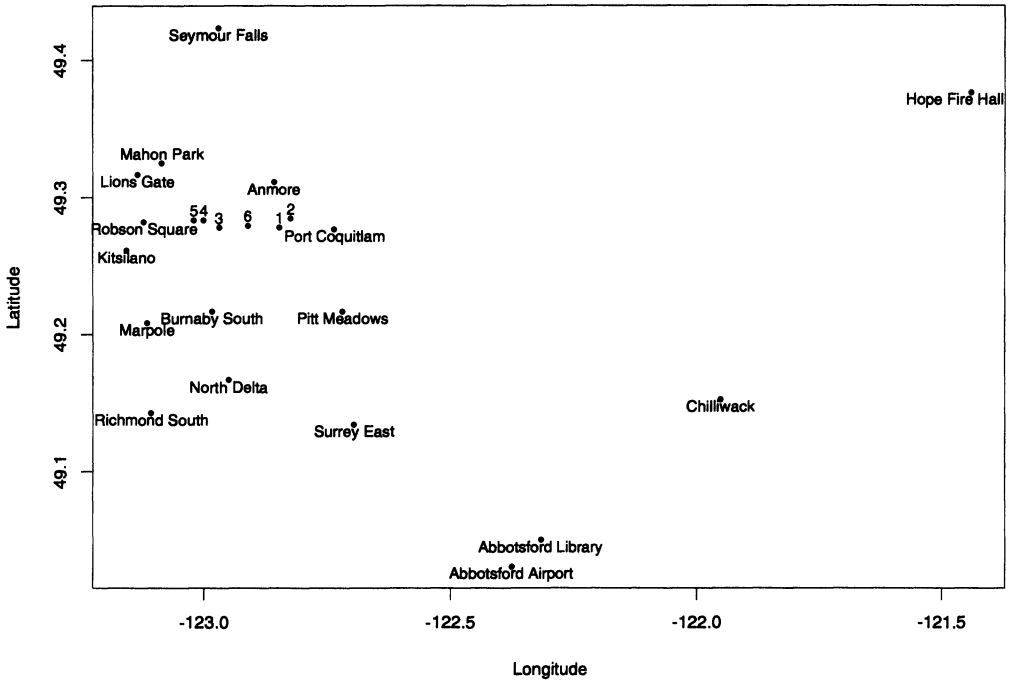
FIGURE 3: Ozone monitor sites (1 - Rocky Point Park; 2 - Eagle Ridge; 3 - Kensington Park; 4 - Confederation Park; 5 - Second Narrows; 6 - Burnaby Mountain).
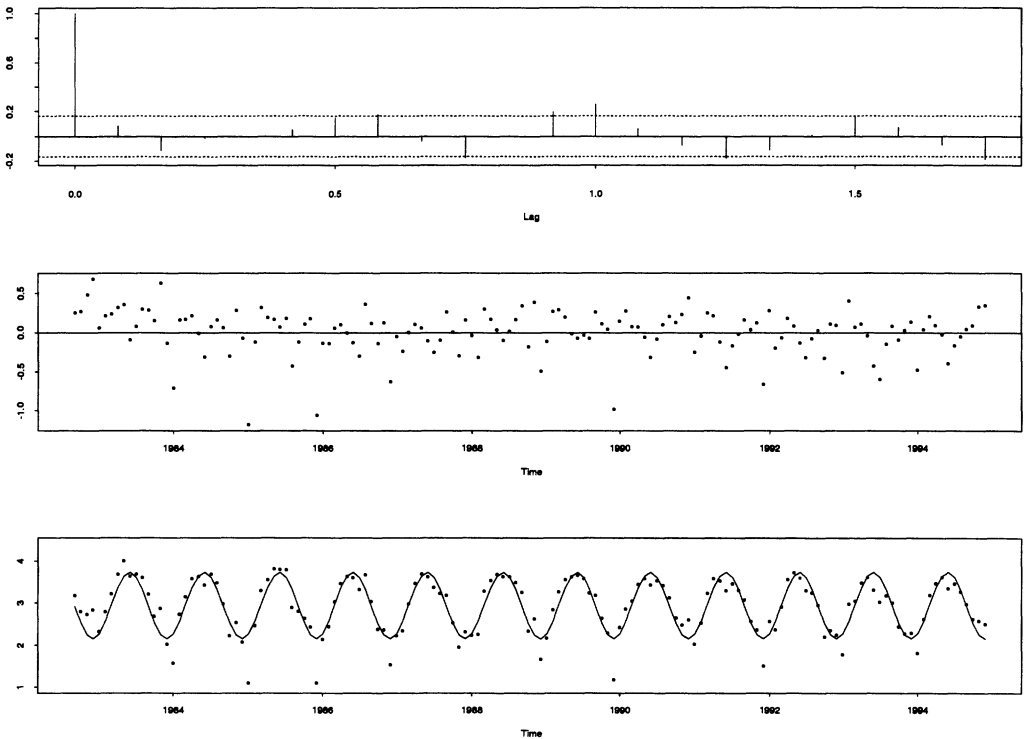

FIGURE 4: Trend modeling at the Rocky Point Park station. Upper: partial autocorrelation function of residuals after removing annual cycle; Middle: residual plot; Lower: fitted trend and observations.

2.2. Covariates and randomly missing data imputation.

In this example, the log-transformed monthly average ozone level is modeled by (1) in Section 3 with covariates $Z = \{1, \cos(2\pi t/12), \sin(2\pi t/12)\}$. This trend model has been used successfully to describe the log-transformed monthly average ozone levels by other authors; see, for example, Brown, Le & Zidek (1994a) for Ontario data.

Let Y_{it} denote the monthly average ozone level at site i for month t. From model (1), given the parameters and covariates, Y_{it} can then be written as

$$Y_{it} = \beta_{i0} + \beta_{i1} \cos\left(\frac{2\pi t}{12}\right) + \beta_{i2} \sin\left(\frac{2\pi t}{12}\right) + \varepsilon_{it},$$

where $(\varepsilon_{1,t}, \ldots, \varepsilon_{23,t})$ are independent over time and follow a normal distribution with mean 0 and variance Σ.

Figure 4 presents the fits of the model to the observed data for a typical site. For brevity, similar results for other sites are omitted. The results indicate a very strong yearly cycle for the data. This pattern was also observed for Ontario data.

Figure 4 also shows the partial autocorrelation function of the fitted residuals. Although an indication of 3-month periodicity is observed in the plot, the partial autocorrelations do not greatly exceed the 5% critical values and so for brevity, they have been ignored in the model. The results therefore suggest that the log-transformed monthly average ozone levels can be assumed to be approximately independent. Also apparent in Figure 4 are low-lying outliers that appear to represent a systematic year-end effect for which we do not have a compelling explanation.

Our theory assumes no randomly missing data within the staircase structure of observed data. Therefore, to proceed with the implementation of the method, we need to impute the randomly missing values, totaling 8% of the data among all stations. These missing observations do not appear to have any obvious pattern. We do the imputation using the standard method for imputation via the EM algorithm (Little & Rubin 1987).

2.3. Spatial interpolation and backcasting.

Using the trend covariates specified above, we apply our model and the results in Theorem 1 to the ozone data where groups of stations with the same starting time are considered as blocks (see Figure 1 for more details).

The hyperparameters are estimated via the EM algorithm as described in Section 4. Given the estimated hyperparameters, the predictive distribution for concentration levels at various locations and different times is fully characterized. This includes the interpolation and backcasting components. We could backcast (i.e., impute the systematically missing values for the monitoring sites, which we will call "gauged" sites) or interpolate (i.e., for ungauged sites) the concentration levels, along with the corresponding predictive intervals. Figure 2 shows the backcast ozone levels and 95% predictive intervals for the Burnaby Mountain station. The predictive intervals are obtained by simulating 1000 realizations at each month from the predictive distribution. This is achieved by sequentially generating realizations from the t-distributions, since the predictive distribution is characterized by a product of matrix-t distributions as given in Theorem 1. Hence predictive intervals can be computed without not much computing power, using subroutines available in standard libraries.

Predictive intervals in Figure 2 for the months between January 1978 and September 1982, when only the first block of stations was in operation, are generally larger than those for the months between September 1982 and July 1986, when the first two blocks were in operation. This is expected since more observed data are available for predictions with two blocks and hence the corresponding predictive intervals are smaller. Backcast levels at other stations (not included here) show similar characteristics.

The predictive distribution for the ungauged sites requires estimates of the hypercovariance matrix among the locations of interest. These are obtained via the powerful method proposed

by Sampson & Guttorp (1992) as described in Section 4.2. Briefly, the Sampson–Guttorp (SG) method first establishes a smooth mapping function between locations in the geographic space (G-space) where the spatial field under consideration is generally not stationary and locations in a new space, called dispersion space (D-space) where the spatial field is isotropic. The mapping function is estimated using the observed correlations between the monitoring sites. A fitted variogram, or equivalently the correlation function, in the D-space and the estimated mapping function are then used to obtain spatial correlations between all locations of interest.

Figure 5 demonstrates the actions of the SG method in this application. The right panel shows the corresponding D-space coordinates, resulting from applying the mapping function to a biorthogonal grid in G-space. The left panel shows the fitted variogram in D-space. The horizontal dotted line there depicts a zero correlation corresponding to a dispersion of 2. The results in the panels can be used to estimate spatial correlations between any points in the G-space, e.g., by first identifying the points in D-space using the grid, then measuring the distance in D-space between them, and finally applying the fitted variogram to the distance to estimate their spatial correlations. Note that the SG method does not require that the units of the D-plane coordinates be explicitly specified. There is a built-in smoothing parameter in the mapping function to control the distortion between the G-space and the D-space. This feature allows users to ensure that the grid is not folded in the D-space and hence to maintain the spatial interpretability of the correlations; locally, then, the closer the stations are together, the higher the correlations.
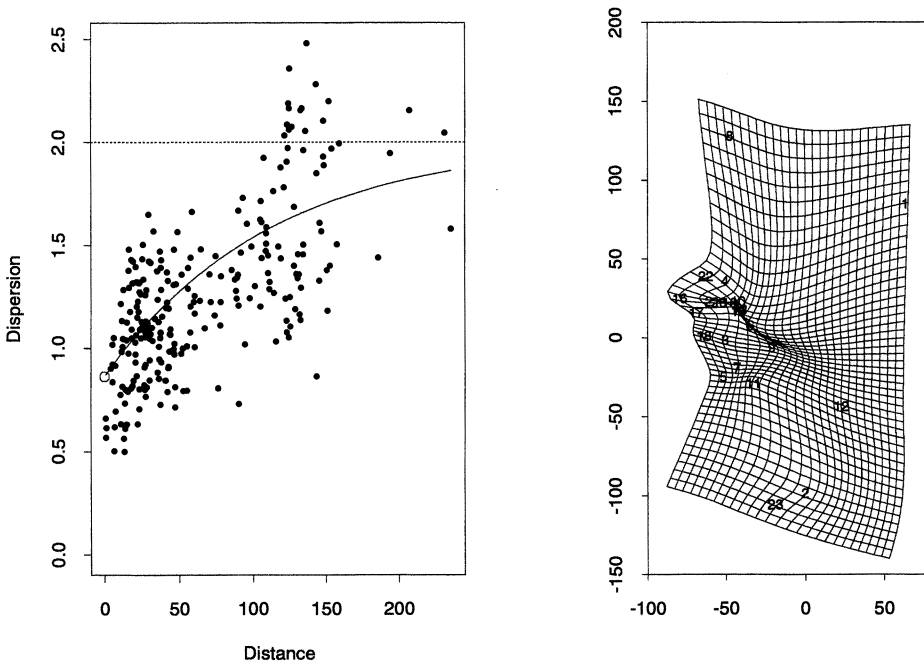
FIGURE 5: Fitted variogram using D-space interstation distance (left panel) and D-space coordinates (right panel) with smoothing parameter equal to 0.3 used in the mapping function.

Figure 6 displays four interpolated monthly average ozone levels (in $\mu g/m^3$) in 1994. The seasonal variation of the monthly average ozone levels is quite obvious, with the lowest level in winter. As expected, the interpolated concentration levels at locations in the vicinity of the monitoring stations are strongly influenced by the observed levels at nearby stations. For locations far from monitoring stations, the interpolated levels are quite close to the overall average of the observed values at all stations.

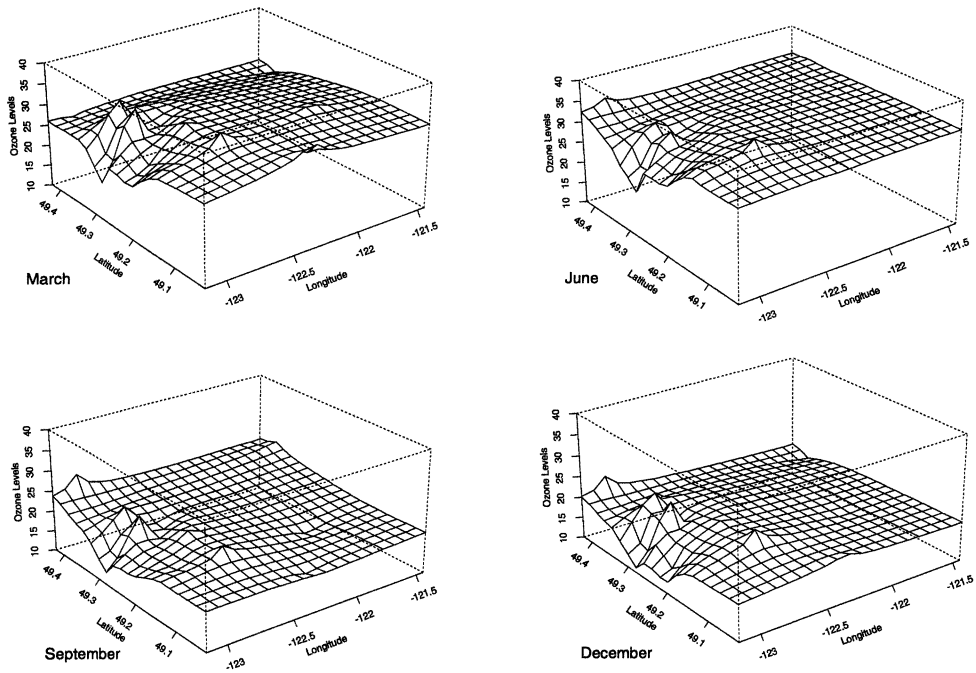FIGURE 6: Interpolated monthly average ozone levels ($\mu g/m^3$) for four months in 1994.

This feature is more clearly demonstrated in Figure 7, which shows the contour plot of the interpolated average ozone level in June 1994. To assess the uncertainty of the interpolator, we present the standard deviation contour plot for the interpolations in June 1994 in Figure 8.

The contour plot shows that the predictive distribution has a smaller standard deviation at locations close to the monitoring sites as expected. Similarly, locations that are far from any monitor, such as the center of the region and the southeast corner, have larger standard deviations.

Overall, the results in this application indicate that the method works well in that the spatial predictions capture not only the temporal features demonstrated at specific stations, but also spatial features of the pollution field. The uncertainty reflected in the predictive distribution seems intuitively appropriate with more observed measurements yielding a smaller standard deviation. In the next section, we describe the main theoretical results.

3. MAIN RESULTS

3.1. Notation.

Throughout the paper, $\otimes$ represents the Kronecker product between matrices. In addition, we let

$$n = \text{number of time points (e.g., number of months)};$$
$$u = \text{number of locations with no monitors — called ungauged sites};$$
$$g = \text{number of locations with monitors — called gauged sites}.$$

The g gauged sites are organized into k blocks such that the jth block consists of g_j stations having the same number m_j of missing responses and hence, $g = g_1 + \cdots + g_k$. These blocks are numbered so that the observed measurements correspond to a monotone data pattern or a staircase structure, i.e.,

$$m_1 \geq m_2 \geq \cdots \geq m_k \geq 0.$$

If the response values prior to the first monitor in operation are of interest, then m_k is set to be bigger than 0.
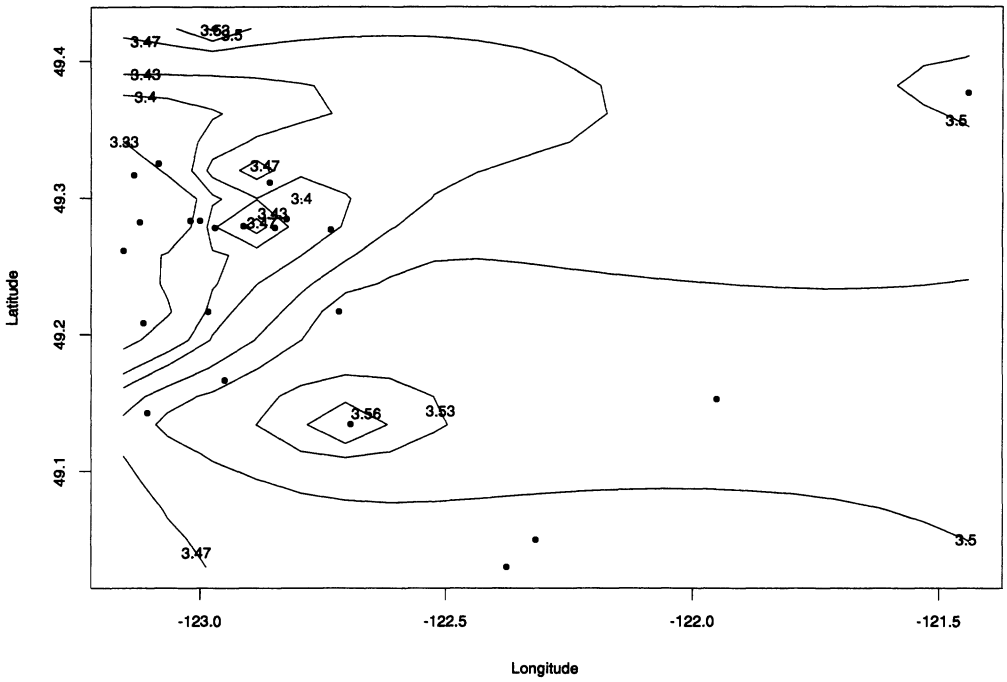
FIGURE 7: Contour plot of the interpolated log ozone field ($\ln \mu g/m^3$) in June, 1994.
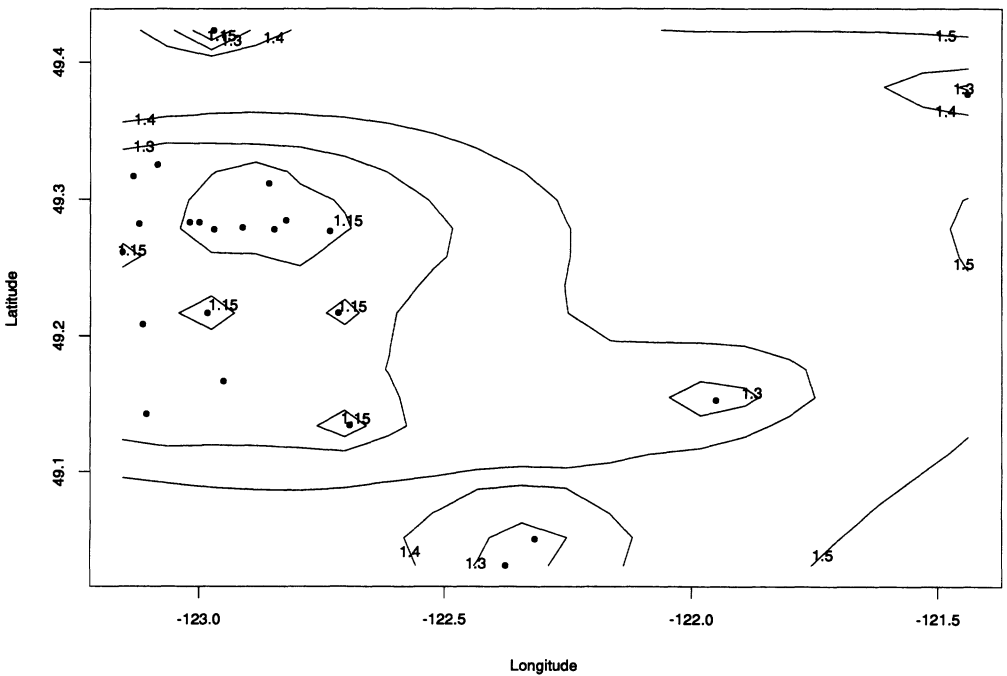The dots denote the stations.

FIGURE 8: Contour plot of the standard deviations for the interpolations
in log-scale ($\ln \mu g/m^3$) in June 1994.

The following notation is used to facilitate our presentation. Denote

(a) the response variables at the gauged and ungauged sites by

$$
Y = \left[Y^{[u]}, \left(\begin{array}{c} Y^{[g_1^m]} \\ Y^{[g_1^o]} \end{array} \right), \ldots, \left(\begin{array}{c} Y^{[g_k^m]} \\ Y^{[g_k^o]} \end{array} \right) \right],
$$

where

  (i) $Y^{[u]}$, $n \times u$ matrix, denotes the responses at ungauged sites,

  (ii) $Y^{[g_j^m]}$, $m_j \times g_j$ matrix, denotes the missing responses at the gauged sites in the jth block,

  (iii) $Y^{[g_j^o]}$, $(n - m_j) \times g_j$ matrix, denotes the observed responses at the gauged sites in the jth block;

(b) the observed measurements at the gauged sites by D where $D = \left\{ Y^{[g_1^o]}, \ldots, Y^{[g_k^o]} \right\}$;

(c) the unobserved responses by $Y_{\text{unob}} = \left\{ Y^{[u]}, Y^{[g_1^m]}, \ldots, Y^{[g_k^m]} \right\}$;

(d) the unobserved responses in blocks j to k by $Y^{[g_j^m, \ldots, g_k^m]} = \left\{ Y^{[g_j^m]}, \ldots, Y^{[g_k^m]} \right\}$;

(e) the responses from blocks j to k, including both observed and unobserved sites by

$$
Y^{[g_j, \ldots, g_k]} = \left[\left(\begin{array}{c} Y^{[g_j^m]} \\ Y^{[g_j^o]} \end{array} \right), \ldots, \left(\begin{array}{c} Y^{[g_k^m]} \\ Y^{[g_k^o]} \end{array} \right) \right];
$$

(f) the responses from all gauged sites by $Y^{[g]} = Y^{[g_1, \ldots, g_k]}$.

We postulate ℓ time-varying covariates responses $Z_t = (Z_{t1}, \ldots, Z_{t\ell})'$, at each time point t being constant across all sites, and write

$$
Z = \left(\begin{array}{c} Z_1' \\ \vdots \\ Z_n' \end{array} \right).
$$

The $\ell \times (u + g)$ coefficient matrix β corresponding to the ℓ covariates and covariance matrix Σ of dimension $(u + g) \times (u + g)$ over gauged and ungauged sites are partitioned conformably:

$$
\beta = \left(\beta^{[u]}, \beta^{[g]} \right), \quad \Sigma = \left(\begin{array}{cc} \Sigma^{[u]} & \Sigma^{[ug]} \\ \Sigma^{[gu]} & \Sigma^{[g]} \end{array} \right).
$$

The coefficient matrix $\beta^{[g]}$ for the gauged sites is further partitioned by blocks as

$$
\beta^{[g]} = \left(\beta^{[g_1]}, \ldots, \beta^{[g_k]} \right) \quad \text{and} \quad \beta^{[g_j, \ldots, g_k]} = \left(\beta^{[g_j]}, \ldots, \beta^{[g_k]} \right).
$$

Correspondingly, we partition $\Sigma^{[g]}$ as

$$
\Sigma^{[g]} = \left(\begin{array}{ccc} \Sigma^{[g_1]} & \ldots & \Sigma^{[g_1, g_k]} \\ \ldots & \ldots & \ldots \\ \Sigma^{[g_k, g_1]} & \ldots & \Sigma^{[g_k]} \end{array} \right), \quad \Sigma^{[g_j, \ldots, g_k]} = \left(\begin{array}{ccc} \Sigma^{[g_j]} & \ldots & \Sigma^{[g_j, g_k]} \\ \ldots & \ldots & \ldots \\ \Sigma^{[g_k, g_j]} & \ldots & \Sigma^{[g_k]} \end{array} \right).
$$

The following one-to-one transformation (Bartlett 1933) of the matrix Σ is used:

$$\Sigma_{kk} = \Sigma^{[g_k]},$$

$$\Gamma_j = \Sigma^{[g_j]} - \Sigma^{[g_j,(g_{j+1},\ldots,g_k)]} \left(\Sigma^{[g_{j+1},\ldots,g_k]}\right)^{-1} \Sigma^{[(g_{j+1},\ldots,g_k),g_j]},$$

$$\tau_j = \left(\Sigma^{[g_{j+1},\ldots,g_k]}\right)^{-1} \Sigma^{[(g_{j+1},\ldots,g_k),g_j]},$$

where

$$\Sigma^{[(g_{j+1},\ldots,g_k),g_j]} = \begin{pmatrix} \Sigma^{[g_{j+1},g_j]} \\ \vdots \\ \Sigma^{[g_k,g_j]} \end{pmatrix} ; \quad j = 1,\ldots,k-1.$$

The matrix $\Sigma^{[g]}$ can then be obtained from $\{\Sigma_{kk}, (\Gamma_{k-1},\tau_{k-1}),\ldots,(\Gamma_1,\tau_1)\}$ by means of this transformation.

3.2. The model.

The response matrix Y is assumed to follow a Gaussian–generalized inverted Wishart model. Specifically, using the notation described above,

$$\begin{cases} Y \mid \boldsymbol{\beta}, \Sigma \sim N(Z\boldsymbol{\beta}, I_n \otimes \Sigma), \\[2mm] \boldsymbol{\beta} \mid \Sigma, \boldsymbol{\beta}_0, F \sim N(\boldsymbol{\beta}_0, F^{-1} \otimes \Sigma), \\[2mm] \Sigma \sim GIW(\Psi, \delta), \end{cases} \tag{1}$$

where $N(\cdot,\cdot)$ denotes the Gaussian distribution, $\boldsymbol{\beta}_0$ is the $\ell \times (g+u)$ hyperparameter mean matrix of $\boldsymbol{\beta}$, F^{-1} is an $\ell \times \ell$ positive definite matrix representing the variance component of $\boldsymbol{\beta}$ between its ℓ rows, and Z is the matrix of covariates. GIW denotes a generalized inverted Wishart distribution of Σ with $\delta = (\delta_0,\ldots,\delta_k)'$ representing degrees of freedom, and Ψ being a collection of hyperparameters defined below. The GIW distribution is recursively defined by

$$\begin{cases} \Sigma^{[g]} \sim GIW\left(\Psi^{[g]}, \delta^{[g]}\right), \\[2mm] \Gamma^{[u]} \sim IW(\Psi_0, \delta_0), \\[2mm] \tau^{[u]} \mid \Gamma^{[u]} \sim N\left(\tau_{0\,u}, H_0 \otimes \Gamma^{[u]}\right), \end{cases}$$

where $\Gamma^{[u]} = \Sigma^{[u|g]} = \Sigma^{[u]} - \Sigma^{[ug]}(\Sigma^{[g]})^{-1}\Sigma^{[gu]}; \quad \tau^{[u]} = (\Sigma^{[g]})^{-1}\Sigma^{[gu]}$. IW denotes the inverted Wishart distribution with hyperparameters (Ψ_0, δ_0); the matrix $\tau_{0\,u}$ is the hyperparameter of $\tau^{[u]}$; and the matrix H_0 is the variance component of τ_u between its rows.

Moreover, the above GIW distribution is defined in a stepwise fashion through $\Sigma^{[g]}$ with $\delta^{[g]} = (\delta_1,\ldots,\delta_g)$ and $\Psi^{[g]}$ being another collection of hyperparameters. The distribution of $\{\Sigma_{kk}, (\Gamma_{k-1},\tau_{k-1}),\ldots,(\Gamma_1,\tau_1)\}$ is defined as follows:

$$\begin{cases} \Sigma_{kk} \sim IW(\Psi_k, \delta_k), \\[2mm] \tau_j \mid \Gamma_j \sim N(\tau_{0j}, H_j \otimes \Gamma_j), \quad j = 1,\ldots,k-1; \\[2mm] \Gamma_j \sim IW(\Psi_j, \delta_j), \quad j = 1,\ldots,k-1. \end{cases} \tag{2}$$

The hyperparameters involved in our Gaussian-GIW model can be written as

$$\mathcal{H} = \{\beta_0, F, \Psi, \delta\},$$

where

$$\Psi = \{\Psi^{[g]}, \Psi_0, \tau_{0\,u}, H_0\}, \quad \Psi^{[g]} = \{\Psi_k, (\Psi_j, H_j, \tau_{0\,j}); j = 1, \ldots, k-1\},$$

and

$$\delta = (\delta_0, \delta^{[g]}), \quad \delta^{[g]} = (\delta_1, \ldots, \delta_g).$$

The dimensions of the parameters in Ψ are as follows

$$\Psi_0 : u \times u, \quad \tau_{0u} : g \times u, \quad H_0 : g \times g, \quad \Psi_k : g_k \times g_k,$$

and for $j = 1, \ldots, k-1$,

$$\Psi_j : g_j \times g_j, \quad \tau_j : (g_{j+1} + \cdots + g_k) \times g_j, \quad H_j : (g_{j+1} + \cdots + g_k) \times (g_{j+1} + \cdots + g_k).$$

In our work, we adopt

$$\tau_{0j} = \Psi^{-1}_{(j+1,j+1)} \Psi_{(j+1),j} \quad \text{and} \quad H_j = \Psi^{-1}_{(j+1,j+1)}. \tag{3}$$

Furthermore, we assume that the degrees of freedom, $\delta_1, \ldots, \delta_k$, follow a gamma prior distribution (Le, Sun & Zidek, 1998) where, for specified α and r,

$$\pi(\delta) \propto (\delta_1 \cdots \delta_k)^{\alpha-1} \exp\{-r(\delta_1 + \cdots + \delta_k)\}.$$

Remarks.

1. The GIW distribution, introduced by Brown, Le & Zidek (1994b), generalizes the IW distribution by allowing different degrees of freedom for a random positive definite matrix. A $p \times p$ positive definite matrix S has an IW distribution, denoted by $IW(B, \delta)$, with δ degrees of freedom if its density function is proportional to

$$|S|^{-(\delta+p+1)/2} \exp\left\{-\tfrac{1}{2} \operatorname{tr}\left(S^{-1}B\right)\right\}.$$

2. The GIW distribution is a conjugate prior for a Gaussian distribution. This prior is very flexible and quite natural to deal with the staircase structure of the observed data. For example, different degrees of freedom for the k blocks can be expressed through the hyperparameter vector δ.

3. The GIW modeling method also gives us considerable latitude in selecting the numbers of blocks in the GIW structure. For example, we could group all sites that started operation at the same time in one block. Alternatively, we could select each site as a block in the stair-case structure.

3.3. The predictive distributions.

This section showcases our principal results, the joint predictive distributions of all unobserved responses. Their means offer point predictors of those responses while the distribution as a whole allows us to assess the uncertainty of those predictors. Furthermore, they allow us to convolve the unknown function with impact distributions so as to incorporate that uncertainty fully in a hierarchical model.

To facilitate the presentation of the main results, we introduce the following notation. Let

$$\begin{pmatrix} \mu^{[j]}_{(1)} \\ \mu^{[j]}_{(2)} \end{pmatrix} : \begin{pmatrix} m_j \times g_j \\ (n - m_j) \times g_j \end{pmatrix} = Z\beta^{[g_j]}_0 + \tilde{\varepsilon}^{[g_{j+1}, \ldots, g_k]} \tau_{0j},$$

$$
\begin{pmatrix} A_{11}^{[j]} & A_{12}^{[j]} \\ A_{21}^{[j]} & A_{22}^{[j]} \end{pmatrix} : \begin{pmatrix} m_j \times m_j & m_j \times (n - m_j) \\ (n - m_j) \times m_j & (n - m_j) \times (n - m_j) \end{pmatrix}
$$

$$
= I_n + Z F^{-1} Z' + \tilde{\varepsilon}^{[g_{j+1},\ldots,g_k]} H_j \left(\tilde{\varepsilon}^{[g_{j+1},\ldots,g_k]} \right)',
$$

where

$$
\tilde{\varepsilon}^{[g_{j+1},\ldots,g_k]} = \begin{cases} Y^{[g_{j+1},\ldots,g_k]} - Z \boldsymbol{\beta}_0^{[g_{j+1},\ldots,g_k]}, \text{ for } j = 1, \ldots, k-1, \\ \\ 0, \text{ for } j = k. \end{cases}
$$

Moreover, for $j = 1, \ldots, k$,

$$
\mu_{(1|2)}^{[j]} = \mu_{(1)}^{[j]} + A_{12}^{[j]} \left(A_{22}^{[j]} \right)^{-1} \tilde{e}^{[g_j^o]},
$$

$$
\Phi_{(1|2)}^{[j]} = \frac{\delta_j - g_j + 1}{\delta_j - g_j + n - m_j + 1} \left\{ A_{11}^{[j]} - A_{12}^{[j]} \left(A_{22}^{[j]} \right)^{-1} A_{21}^{[j]} \right\},
$$

$$
\Psi_{(1|2)}^{[j]} = \frac{1}{\delta_j - g_j + 1} \left\{ \Psi_j + \left(\tilde{e}^{[g_j^o]} \right)' \left(A_{22}^{[j]} \right)^{-1} \tilde{e}^{[g_j^o]} \right\},
$$

$$
\delta_{(1|2)}^{[j]} = \delta_j - g_j + n - m_j + 1,
$$

where $\tilde{e}^{[g_j^o]} = Y^{[g_j^o]} - \mu_{(2)}^{[j]}$,

$$
\mu^{[u|g]} = Z \boldsymbol{\beta}_0^{[u]} + \tilde{\varepsilon}^{[g]} \tau_{0u}, \quad \Phi^{[u|g]} = I_n + Z F^{-1} Z' + \tilde{\varepsilon}^{[g]} H_0 \left(\tilde{\varepsilon}^{[g]} \right)',
$$

and $\tilde{\varepsilon}^{[g]} = Y^{[g]} - Z \boldsymbol{\beta}_0^{[g]}$.

A matrix-valued random variable $X_{n \times m}$ is said to have a matrix t-distribution, i.e.,

$$
X \sim t_{n \times m} \left(X^{(0)}, A \otimes B, \delta \right),
$$

where A is $n \times n$ and B is $m \times m$, if its density function has the form

$$
f(X) \propto |A|^{-m/2} |B|^{-n/2} \left| I_n + \delta^{-1} \left\{ A^{-1} (X - X^{(0)}) \right\} \left\{ (X - X^{(0)}) B^{-1} \right\}' \right|^{-(\delta + n + m - 1)/2},
$$

and the normalizing constant of the density is given by

$$
K = (\delta \pi^2)^{-\frac{nm}{2}} \frac{\Gamma_{n+m} \{ (\delta + n + m - 1)/2 \}}{\Gamma_n \{ (\delta + n - 1)/2 \} \Gamma_m \{ (\delta + m - 1)/2 \}},
$$

where

$$
\Gamma_p(t) = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma \{ t - (i - 1)/2 \}
$$

denotes the multivariate Gamma function.

THEOREM 1. *The predictive distribution of the unobserved responses conditional on the observed data D and the hyperparameter set $\mathcal{H}$ is given by*

$$(Y_{\text{unob}} \mid D, \mathcal{H}) \sim \left(Y^{[u]} \mid Y^{[g_1^m, \ldots, g_k^m]}, D, \mathcal{H} \right) \prod_{j=1}^{k-1} \left(Y^{[g_j^m]} \mid Y^{[g_{j+1}^m, \ldots, g_k^m]}, D, \mathcal{H} \right) \times \left(Y^{[g_k^m]} \mid D, \mathcal{H} \right),$$

(4)

where the three components of the conditional distributions are specified as follows:

$$\left(Y^{[g_k^m]} \mid D, \mathcal{H} \right) \sim t_{m_k \times g_k} \left(\mu_{(1|2)}^{[k]}, \Phi_{(1|2)}^{[k]} \otimes \Psi_{(1|2)}^{[k]}, \delta_{(1|2)}^{[k]} \right),$$

(5)

$$\left(Y^{[g_j^m]} \mid Y^{[g_{j+1}^m, \ldots, g_k^m]}, D, \mathcal{H} \right) \sim t_{m_j \times g_j} \left(\mu_{(1|2)}^{[j]}, \Phi_{(1|2)}^{[j]} \otimes \Psi_{(1|2)}^{[j]}, \delta_{(1|2)}^{[j]} \right),$$

(6)

$$\left(Y^{[u]} \mid Y^{[g_1^m, \ldots, g_k^m]}, D, \mathcal{H} \right) \sim t_{n \times u} \left(\mu^{[u|g]}, (\delta_0 - u + 1)^{-1} \Phi^{[u|g]} \otimes \Psi_0, \delta_0 - u + 1 \right).$$

(7)

Remarks.

1. We describe (5) and (6) as backcasting since they give the joint predictive distribution of the response variables at the gauged sites during their ungauged time period. More precisely,

$$(Y_{\text{backcasting}} \mid D, \mathcal{H}) = \prod_{j=1}^{k-1} \left(Y^{[g_j^m]} \mid Y^{[g_{j+1}^m, \ldots, g_k^m]}, D, \mathcal{H} \right) \times \left(Y^{[g_k^m]} \mid D, \mathcal{H} \right).$$

2. We describe (7) as spatial interpolation since it is the predictive distribution of the response variables at the ungauged sites during the time period under consideration. More precisely,

$$(Y_{\text{interpolation}} \mid Y_{\text{backcasting}}, D, \mathcal{H}) = \left(Y^{[u]} \mid Y^{[g_1^m, \ldots, g_k^m]}, D, \mathcal{H} \right).$$

3. The result (4) can be used to obtain predictive distributions for forecasting. This can be achieved by appropriately choosing Z corresponding to the first m_k components. For example, to forecast the $(n+1)$th month in the above application, we let the first component of Z be $[1, \cos\{2\pi(n+1)/12\}, \sin\{2\pi(n+1)/12\}]$.

3.4. The posterior distributions.

While our interest in this paper centers on predicting unobserved responses, we recognize that in some applications the model parameters themselves will be of inferential interest. The model "transfer" coefficients in β, for example, give insight into the role of the covariates in shaping the joint response surface. The spatial covariance may reveal the influence of latent factors such as wind speed or direction. Thus we develop and investigate the relevant posterior distributions in this section.

To present those posterior distributions, we adopt the following basic notation:

$$K_j = (0, I_{n-m_j}) : (n - m_j) \times n, \quad Z_{(j)} = K_j Z,$$

$$Y_{(j)}^{[g_{j+1}, \ldots, g_k]} = K_j Y^{[g_{j+1}, \ldots, g_k]}, \quad \tilde{F}_j = Z_{(j)} Z_{(j)}' + F, \quad W_j = \tilde{F}_j^{-1} Z_{(j)}' Z_{(j)},$$

for $j = 1, \ldots, k$. Next we define notation that uses the basic notation:

$$
\begin{aligned}
\hat{\beta}^{[g_j^\circ]} &= (Z_{(j)} Z_{(j)}')^{-1} Z_{(j)}' Y^{[g_j^\circ]}, \\
\hat{\beta}^{[g_{j+1}, \ldots, g_k]} &= (Z_{(j)} Z_{(j)}')^{-1} Z_{(j)}' Y_{(j)}^{[g_{j+1}, \ldots, g_k]}, \\
\tilde{\beta}^{[g_j^\circ]} &= W_j \hat{\beta}^{[g_j^\circ]} + (I - W_j) \beta_0^{[g_j]}, \\
\tilde{\beta}^{[g_{j+1}, \ldots, g_k]} &= W_j \hat{\beta}^{[g_{j+1}, \ldots, g_k]} + (I - W_j) \beta_0^{[g_{j+1}, \ldots, g_k]}, \\
\tilde{\varepsilon}^{[g_j^\circ]} &= Y^{[g_j^\circ]} - Z_{(j)} \beta_0^{[g_j]}, \\
\tilde{\varepsilon}_{(j)}^{[g_{j+1}, \ldots, g_k]} &= Y_{(j)}^{[g_{j+1}, \ldots, g_k]} - Z_{(j)} \beta_0^{[g_{j+1}, \ldots, g_k]}.
\end{aligned}
$$

Finally, let

$$
\begin{aligned}
\tilde{\Psi}_k &= \Psi_k + (\tilde{\varepsilon}^{[g_k^\circ]})' \left(I_{n-m_k} + Z_{(k)} F^{-1} Z_{(k)}' \right)^{-1} \tilde{\varepsilon}^{[g_k^\circ]}, \\
\tilde{\Psi}_j &= \Psi_j + \left(\tilde{\varepsilon}^{[g_j^\circ]} - \tilde{\varepsilon}_{(j)}^{[g_{j+1}, \ldots, g_k]} \tau_{0j} \right)' \left\{ I_{n-m_j} + Z_{(j)} F^{-1} Z_{(j)}' \right. \\
&\quad \left. + \left(\tilde{\varepsilon}_{(j)}^{[g_{j+1}, \ldots, g_k]} \right)' H_j \tilde{\varepsilon}_{(j)}^{[g_{j+1}, \ldots, g_k]} \right\} \left(\tilde{\varepsilon}^{[g_j^\circ]} - \tilde{\varepsilon}_{(j)}^{[g_{j+1}, \ldots, g_k]} \tau_{0j} \right), \\
\tilde{H}_j^{-1} &= H_j^{-1} + \left(\tilde{\varepsilon}_{(j)}^{[g_{j+1}, \ldots, g_k]} \right)' \left(I_{n-m_j} + Z_{(j)} F^{-1} Z_{(j)}' \right)^{-1} \tilde{\varepsilon}_{(j)}^{[g_{j+1}, \ldots, g_k]}, \\
\tilde{\tau}_j &= \tilde{H}_j \left\{ H_j^{-1} \tau_{0j} + \left(\tilde{\varepsilon}_{(j)}^{[g_{j+1}, \ldots, g_k]} \right)' \left(I_{n-m_j} + Z_{(j)} F^{-1} Z_{(j)}' \right)^{-1} \tilde{\varepsilon}^{[g_j^\circ]} \right\}, \\
\tilde{\delta}_j &= \delta_j + n - m_j,
\end{aligned}
$$

for $j = 1, \ldots, k-1$. We can now state the main theorem of this section.

THEOREM 2. *The joint posterior density is*

$$
f(\beta, \Sigma \mid D, \mathcal{H}) = f(\beta \mid \Sigma, D, \mathcal{H}) f(\Sigma \mid D, \mathcal{H}),
$$

where

$$
f(\beta \mid \Sigma, D, \mathcal{H}) = f(\beta^{[g_k]} \mid D, \Sigma_{kk}, \mathcal{H}) \prod_{j=1}^{k-1} f(\beta^{[g_j]} \mid D, \beta^{[g_{j+1}, \ldots, g_k]}, \tau_j, \Gamma_j, \mathcal{H})
$$

with

$$
\beta^{[g_k]} \mid D, \Sigma_{kk}, \mathcal{H} \sim N \left(\tilde{\beta}^{[g_k^\circ]}, \tilde{F}_k^{-1} \otimes \Sigma_{kk} \right),
$$

$$
\beta^{[g_j]} \mid D, \beta^{[g_{j+1}, \ldots, g_k]}, \tau_j, \Gamma_j, \mathcal{H} \sim N \left(\tilde{\beta}^{[g_j^\circ]} + \left(\beta^{[g_{j+1}, \ldots, g_k]} - \tilde{\beta}^{[g_{j+1}, \ldots, g_k]} \right) \tau_j, \tilde{F}_j^{-1} \otimes \Gamma_j \right),
$$

$$\tag{8}$$

and

$$
f(\Sigma \mid D, \mathcal{H} = f(\Sigma_{kk} \mid D, \mathcal{H}) \prod_{j=1}^{k-1} f(\tau_j \mid D, \Gamma_j, \mathcal{H}) f(\Gamma_j \mid D, \mathcal{H})
$$

with

$$
\Sigma_{kk} \mid D, \mathcal{H} \sim IW(\tilde{\Psi}_k, \tilde{\delta}_k), \quad \tau_j \mid D, \Gamma_j, \mathcal{H} \sim N \left(\tilde{\tau}_j, \tilde{H}_j \otimes \Gamma_j \right), \quad \Gamma_j \mid D, \mathcal{H} \sim IW(\tilde{\Psi}_j, \tilde{\delta}_j).
$$

$$\tag{9}$$

COROLLARY 1. *The posterior means of β given data D and hyperparameter set $\mathcal{H}$ are given by*

$$\mathrm{E}\left(\beta^{[g_k]} \mid D, \mathcal{H}\right) = \tilde{\beta}^{[g_k^o]},$$

$$\mathrm{E}\left(\beta^{[g_j]} \mid D, \mathcal{H}\right) = \tilde{\beta}^{[g_j^o]} + \left\{\mathrm{E}\left(\beta^{[g_{j+1},\ldots,g_k]} \mid D, \mathcal{H}\right) - \tilde{\beta}^{[g_{j+1},\ldots,g_k]}\right\}\tilde{\tau}_j,$$

where

$$\mathrm{E}\left(\beta^{[g_{j+1},\ldots,g_k]} \mid D, \mathcal{H}\right) = \left\{\mathrm{E}\left(\beta^{[g_{j+1}]} \mid D, \mathcal{H}\right), \ldots, \mathrm{E}\left(\beta^{[g_k]} \mid D, \mathcal{H}\right)\right\}$$

is computed recursively for $j = 1, \ldots, k-1$.

We note that $\mathrm{E}(\beta^{[g_j]} \mid D, \mathcal{H}) = \tilde{\beta}^{[g_j^o]}$ for $j = 1, \ldots, k-1$, when the data are complete, i.e., when D contains no missing blocks.

The posterior means of $(\Sigma_{kk}, \tau_j, \Gamma_j, j = 1, \ldots, k-1)$ can be obtained directly from the Gaussian and inverted Wishart distributions using the theorem above. Other types of posterior expectations relevant to the estimation of parameters are given in Appendix B.

4. PARAMETER ESTIMATION

In this section, we discuss how the elements of the hyperparameter set $\mathcal{H}$ are estimated. These estimates are obtained in two steps. In the first, where data are available for direct estimation, the hyperparameters are found using the type II maximum likelihood approach. The second step involves the estimation of the hyperparameters associated with the ungauged sites. This step is done using the spatial covariance interpolator developed by Sampson & Guttorp (1992).

4.1. Estimating hyperparameters using the EM algorithm.

In the first step of the estimation procedure, we derive the EM algorithm corresponding to the model developed in the previous sections for estimating the hyperparameters. The resulting estimates are the Type II maximum likelihood estimates. This approach has been used in previous studies (Brown, Le & Zidek 1994a; Sun, Le, Zidek & Burnett 1998).

Recall that the hyperparameters $(\Psi_k, [\Psi_j, \tau_{0j}, H_j], j = 1, \ldots, k-1)$ correspond to the partition of Σ through the Bartlett transformation; i.e., with the recursive notation

$$\Psi_{(jj)} = \begin{pmatrix} \Psi_{jj} & \Psi_{j(j+1)} \\ \Psi_{(j+1)j} & \Psi_{(j+1,j+1)} \end{pmatrix}, \quad \text{for } j = 1, \ldots, k-1,$$

the parameterization of $(\Psi_k, [\Psi_j, \tau_{0j}, H_j], j = 1, \ldots, k-1)$ satisfies

$$\Psi_k = \Psi_{kk}, \quad \Psi_j = \Psi_{jj} - \Psi_{j(j+1)}\Psi_{(j+1,j+1)}^{-1}\Psi_{(j+1)j}.$$

It is easy to show from (3) that

$$\Psi_{(jj)} = \begin{pmatrix} \Psi_j + \tau_{0j}'\Psi_{(j+1,j+1)}\tau_{0j} & \tau_{0j}'\Psi_{(j+1,j+1)} \\ \Psi_{(j+1,j+1)}\tau_{0j} & \Psi_{(j+1,j+1)} \end{pmatrix}$$

and

$$|\Psi_{(jj)}| = |\Psi_j| \cdots |\Psi_k|.$$

The EM iterative algorithm (Dempster, Laird & Rubin 1977; Chen 1979; Brown, Le & Zidek 1994a) requires at iteration $p+1$, in the "E-step" the computation of

$$\begin{aligned} Q(\mathcal{H} \mid \mathcal{H}^{(p)}) &= \mathrm{E}\left[\log\{f(Y, \beta, \Sigma \mid \mathcal{H})\}\pi(\delta) \mid D, \mathcal{H}^{(p)}\right] \\ &= \mathrm{E}\left[\log f\{Y \mid \beta, \Sigma\} \mid D, \mathcal{H}^{(p)}\right] + \mathrm{E}\left\{\log f(\beta, \Sigma \mid \mathcal{H}) \mid D, \mathcal{H}^{(p)}\right\} + \log \pi(\delta) \end{aligned}$$

given the previous parameter estimate $\mathcal{H}^{(p)}$ from iteration p. Then at the "M-step" we are required to maximize $Q(\mathcal{H} \mid \mathcal{H}^{(p)})$ over $\mathcal{H}$ to get $\mathcal{H}^{(p+1)}$. Here, the expectation is taken over β and Σ with respect to the posterior distribution $\beta, \Sigma \mid D, \mathcal{H}^{(p)}$.

Notice that $\mathrm{E}\{\log f(Y \mid \beta, \Sigma) \mid D, \mathcal{H}^{(p)}\}$ does not depend on $\mathcal{H}$. Thus the algorithm requires only that we compute

$$Q^*(\mathcal{H} \mid \mathcal{H}^{(p)}) = \mathrm{E}\left\{\log f(\beta, \Sigma \mid \mathcal{H}) \mid D, \mathcal{H}^{(p)}\right\} + \log \pi(\delta)$$

at the E-step and maximize Q^* over $\mathcal{H}$ at the M-step.

With the parameterization introduced above

$$f(\beta, \Sigma \mid \mathcal{H}) \propto \left\{\prod_{j=1}^{k} c(g_j, \delta_j)\right\} |\Sigma_{kk}|^{-\frac{l+\delta_k+g_k+1}{2}} \left(\prod_{j=1}^{k-1} |\Gamma_j|^{-\frac{l+\delta_j+g_j+\cdots+g_k+1}{2}}\right)$$

$$\times \left(\prod_{j=1}^{k} |\Psi_j|^{\frac{\delta_j+g_1+\cdots+g_{j-1}}{2}}\right) |F|^{\frac{g}{2}} \exp\left(-\tfrac{1}{2}\mathrm{tr}\left[\Sigma^{-1}\{\Psi + (\beta - \beta_0)'F(\beta - \beta_0)\}\right]\right),$$

where $g_0 = -g_1$ and

$$c(p, \delta) = \left\{2^{\delta p/2}\Gamma_p(\delta/2)\right\}^{-1}.$$

Hence

$$
\begin{aligned}
Q^*(\mathcal{H} \mid \mathcal{H}^{(p)}) &= \mathrm{E}\left\{\log f(\beta, \Sigma \mid \mathcal{H}) \mid D, \mathcal{H}^{(p)}\right\} + \log \pi(\delta) \\[1ex]
&= \mathrm{constant} + \sum_{j=1}^{k} \log c(g_j, \delta_j) + \log \pi(\delta) \\[1ex]
&\quad -\frac{1}{2}(\ell + \delta_k + g_k + 1)\mathrm{E}\left(\log |\Sigma_{kk}| \mid D, \mathcal{H}^{(p)}\right) \\[1ex]
&\quad -\frac{1}{2}\sum_{j=1}^{k-1}(\ell + \delta_j + g_j + \cdots + g_k + 1)\mathrm{E}\left(\log |\Gamma_j| \mid D, \mathcal{H}^{(p)}\right) \\[1ex]
&\quad +\frac{1}{2}\sum_{j=1}^{k}(\delta_j + g_1 + \cdots + g_{j-1})\log |\Psi_j| + \frac{g}{2}\log |F| \\[1ex]
&\quad -\frac{1}{2}\mathrm{tr}\,\mathrm{E}\left[\Sigma^{-1}\{\Psi + (\beta - \beta_0)'F(\beta - \beta_0)\} \mid D, \mathcal{H}^{(p)}\right].
\end{aligned}
\tag{10}
$$

Suppose the current estimate is

$$\mathcal{H}^{(p)} = \left(\beta_0^{(p)}, F^{(p)}, \Psi_k^{(p)}, \delta_k^{(p)}, [\tau_{0j}^{(p)}, \Psi_j^{(p)}, \delta_j^{(p)}], j = 1, \ldots, k-1\right).$$

The EM algorithm at step $(p+1)$ is implemented in two steps.

(i) E-step: Compute the posterior expectations involved in (10), given data and $\mathcal{H}^{(p)}$. We present the detailed calculations in the Appendix B.

(ii) M-step: Maximize $Q^*(\mathcal{H} \mid \mathcal{H}^{(p)})$ over $\mathcal{H}$ to obtain the updated estimate $\mathcal{H}^{(p+1)}$ of $\mathcal{H}$ at step $(p+1)$. This M-step is carried out by the following updating processes:

(a) To update the estimates of δ, maximize

$$\sum_{j=1}^{k} \log c(g_j, \delta_j) + \log \pi(\delta) - \frac{1}{2}\sum_{j=1}^{k} \delta_j \left\{\mathrm{E}\left(\log |\Gamma_j| \mid D, \mathcal{H}^{(p)}\right) - \log |\Psi_j|\right\},$$

where we have set $\Gamma_k = \Sigma_{kk}$ for simplicity.

(b) To update the estimates of Ψ, maximize

$$\sum_{j=1}^{k} (\delta_j + g_1 + \cdots + g_{j-1}) \log |\Psi_j| - \mathrm{tr} \left\{ \Psi \mathrm{E} \left(\Sigma^{-1} \mid D, \mathcal{H}^{(p)} \right) \right\}.$$

(c) To update the estimates of β_0 and F, maximize

$$g \log |F| - \mathrm{tr} \left[F \mathrm{E} \left\{ (\beta - \beta_0) \Sigma^{-1} (\beta - \beta_0)' \mid D, \mathcal{H}^{(p)} \right\} \right].$$

Note that the estimation of (δ, Ψ) can be separated from that of (β_0, F).
To complete (a)–(c), let us consider the following recursive partition of

$$\mathrm{E} \left(\Sigma^{-1} \mid D, \mathcal{H}^{(p)} \right)$$

$$= \begin{pmatrix} ccB_{11}^{(1)} & B_{12}^{(1)} \\ B_{21}^{(1)} & B_{22}^{(1)} \end{pmatrix} : \begin{pmatrix} g_1 \times g_1 & g_1 \times (g_2 + \cdots + g_k) \\ (g_2 + \cdots + g_k) \times g_1 & (g_2 + \cdots + g_k) \times (g_2 + \cdots + g_k) \end{pmatrix}$$

and

$$B_{22}^{(j-1)} - B_{21}^{(j-1)} \left(B_{11}^{(j-1)} \right)^{-1} B_{12}^{(j-1)}$$

$$= \begin{pmatrix} B_{11}^{(j)} & B_{12}^{(j)} \\ B_{21}^{(j)} & B_{22}^{(j)} \end{pmatrix} : \begin{pmatrix} g_j \times g_j & g_j \times (g_{j+1} + \cdots + g_k) \\ (g_{j+1} + \cdots + g_k) \times g_j & (g_{j+1} + \cdots + g_k) \times (g_{j+1} + \cdots + g_k) \end{pmatrix}$$

for $j = 2, \ldots, k - 1$. Note that the partitions depend also on p, which is dropped from the index for simplicity. Let $\psi(x) = d\{\log \Gamma(x)\}/dx$ denote the digamma function. It follows that the estimates of δ and Ψ can be updated by solving the equations

$$\sum_{i=1}^{g_j} \left\{ \psi \left(\frac{\tilde{\delta}_j^{(p)} - i + 1}{2} \right) - \psi \left(\frac{\delta_j^{(p+1)} - i + 1}{2} \right) \right\} + \log \left| \Psi_j^{(p+1)} \right| - \log \left| \tilde{\Psi}_j^{(p)} \right| = 2r - \frac{2(\alpha - 1)}{\delta_j}$$

for $j = 1, \ldots, k - 1$, then calculating

$$\Psi_j^{(p+1)} = \left(\delta_j^{(p+1)} + g_1 + \cdots + g_{j-1} \right) \left(B_{11}^{(j)} \right)^{-1} \quad \text{and} \quad \tau_{0j}^{(p+1)} = -B_{21}^{(j)} \left(B_{11}^{(j)} \right)^{-1}$$

for $j = 1, \ldots, k - 1$, and

$$\Psi_k^{(p+1)} = B_{22}^{(k-1)} - B_{21}^{(k-1)} \left(B_{11}^{(k-1)} \right)^{-1} B_{12}^{(k-1)},$$

(recall that $g_1 + g_0 = 0$), where $\tilde{\delta}_j^{(p)}$ and $\tilde{\Psi}_j^{(p)}$ are as given in the notation in 3.4 (preceding Theorem 2) with parameters replaced by the estimates at iteration p.
The updated version of estimates of β and F are, respectively,

$$\beta_0^{(p+1)} = \left\{ \mathrm{E} \left(\Sigma^{-1} \mid D, \mathcal{H}^{(p)} \right) \right\}^{-1} \mathrm{E} \left(\Sigma^{-1} \beta \mid D, \mathcal{H}^{(p)} \right)$$

and

$$F^{(p+1)} = g^{-1} \left[\mathrm{E} \left(\beta \Sigma^{-1} \beta' \mid D, \mathcal{H}^{(p)} \right) \right.$$
$$\left. - \mathrm{E} \left(\beta \Sigma^{-1} \mid D, \mathcal{H}^{(p)} \right) \left\{ \mathrm{E} \left(\Sigma^{-1} \mid D, \mathcal{H}^{(p)} \right) \right\}^{-1} \mathrm{E} \left(\Sigma^{-1} \beta' \mid D, \mathcal{H}^{(p)} \right) \right].$$

Iterating these EM steps until convergence produces estimates for the hyperparameters including $\{\Psi_k, (\Psi_j, \tau_{0j}, H_j), j = 1, \ldots, k - 1\}$. These estimated hyperparameters can be used to form an estimate for $\Psi^{[g]}$, associated with the spatial hypercovariance of the gauged sites, through the Bartlett transformation. This estimate is relevant to the estimation of $\Psi^{[u]}$ and $\tau_0^{[u]}$, the hyperparameters corresponding to the ungauged sites via the Sampson–Guttorp method as described below.

4.2. Extending the covariance matrix estimate via the Sampson–Guttorp method.

In this section, we describe how the Sampson–Guttorp method can be used to estimate $\Psi^{[u]}$ and $\tau_0^{[u]}$. Note that the SG method is designed to extend the spatial covariance from the gauged to the ungauged sites. The method and its applications have been described in other publications (Sampson & Guttorp 1992; Brown, Le & Zidek 1994a; Sun, Le, Zidek & Burnett 1998) and hence are not included here.

For this problem, the estimates for $\Psi^{[u]}$ and $\tau_0^{[u]}$ can be obtained by first extending the estimated $\Psi^{[g]}$ to the ungauged sites using the SG method and then converting the resulting spatial hypercovariance to estimate the parameters of interest. To illustrate this step, we let $k = 1$ for simplicity, i.e., assume only one block of gauged sites.

Let M be the hypercovariance of Y. Then

$$M = \begin{pmatrix} M_{uu} & M_{ug} M_{gu} & M_{gg} \end{pmatrix}.$$

Thus,

$$\tau_0^{[u]} = M_{gg}^{-1} M_{gu} \tag{11}$$

and

$$\Psi^{[u]} = \frac{(\delta_1 - g - 1)(\delta^{[u]} - u - 1)}{\delta^{[u]} - 1} \left(M_{uu} - M_{ug} M_{gg}^{-1} M_{gu} \right). \tag{12}$$

The hypercovariance M_{gg} is estimated by

$$\hat{M}_{gg} = \frac{\hat{\Psi}^{[g]}}{\hat{\delta}_1 - g - 1}.$$

The SG method is then used to extend $\hat{M}_{gg}$ to estimate M_{uu}, M_{gu} and M_{ug}. The parameters of interest are estimated through equations (11) and (12) using these estimated M-matrices. More details on the SG method are given in the Application section.

5. CONCLUDING REMARKS

In this work, we have developed a Bayesian approach for spatial and temporal interpolation where observed data from monitoring stations following a staircase structure can be incorporated. The approach is an extension of the Bayesian methodology for spatial interpolation developed earlier by Le & Zidek (1992), which was further extended to a multivariate setting by Brown, Le & Zidek (1994a) and Le, Sun & Zidek (1997). This approach would allow for combining data from active networks with different starting times of operation in the interpolation and thus, providing the desirable property of borrowing-from-strength. The approach appears to work reasonably well in the interpolation of ozone concentration levels in British Columbia. This is not too surprising since the earlier developments of the Bayesian approach seem to work quite well with both real data and in a cross-validation setting (Sun, Le, Zidek & Burnett 1998; Sun 1998).

In this work, the Sampson–Guttorp method is used to estimate the non-stationary spatial covariance field. The estimated spatial correlations may be subject to considerable errors which are not incorporated in the current methodology. Assessments of earlier developments of this approach, however, indicate that predictive intervals are well-calibrated without taking into account that uncertainty (Sun, Le, Zidek & Burnett 1998; Sun 1998).

The development of the interpolation approach comes with a related Bayesian design theory (Le & Zidek 1994; Zidek, Sun & Le 2000). Here it is possible to redesign monitoring networks with such staircase structures. Results for this work will be forthcoming.

It would be desirable to extend this methodology to a multivariate setting where observed data for several pollutants with staircase structure are available for interpolation. This would allow for even more borrowing-from-strength. The work by Sun, Le, Zidek & Burnett (1998) has demonstrated that substantial improvements could be achieved when observed data for different pollutants are used to impute any specific pollutant, particularly for those with high correlations. Work in this direction is currently underway.

We have seen above that monthly averages of ozone concentration levels exhibit only moderate temporal correlations. In contrast, hourly or daily levels of various pollutants including ozone and PM_{10} have strong temporal correlations (Li, Le, Sun & Zidek 1999). Spatial interpolation for these hourly or daily fields is crucial for examining the acute health impact of air pollution. Furthermore, monitoring networks often add or remove stations during their operational periods, creating data patterns that are not considered in this work. Extension of the current method to these aspects is under consideration.

In recent years, several authors have improved the traditional kriging method by incorporating it into various Bayesian frameworks (see, e.g., Handcock & Stein 1993; Handcock & Wallis 1994; De Oliveira, Benjamin & Short 1997; Gaudard, Karson, Linder & Sinha 1999). It would be worth comparing the performances of the proposed method with these Bayesian kriging approaches.

APPENDIX A: LEMMAS AND PROOFS OF THEOREMS

LEMMA 1. *Define matrices* $Y : n \times g$, $\boldsymbol{\beta}$, $\boldsymbol{\beta}_0 : l \times g$, $\Sigma > 0, \Psi > 0 : g \times g$, $Z : n \times l$, $F > 0 : l \times l$, *and* $A > 0 : n \times n$. *The Gaussian-Inverted-Wishart model*

$$
\begin{cases}
Y \mid \boldsymbol{\beta}, \Sigma \sim N(Z\boldsymbol{\beta}, A \otimes \Sigma) \\[2mm]
\boldsymbol{\beta} \mid \Sigma \sim N(\boldsymbol{\beta}_0, F^{-1} \otimes \Sigma) \\[2mm]
\Sigma \sim IW(\Psi, \delta)
\end{cases}
$$

implies the following predictive distribution

$$
Y \sim t_{n \times g} \left[Z\boldsymbol{\beta}_0, (\delta - g + 1)^{-1}(A + ZF^{-1}Z') \otimes \Psi, \delta - g + 1 \right]
$$

and the posterior distributions

$$
\boldsymbol{\beta} \mid \Sigma, Y \sim N(W\hat{\boldsymbol{\beta}} + (I - W)\boldsymbol{\beta}_0, \tilde{F}^{-1} \otimes \Sigma),
$$

$$
\Sigma \mid Y \sim IW(\Psi + (Y - Z\boldsymbol{\beta}_0)'(A + ZF^{-1}Z')^{-1}(Y - Z\boldsymbol{\beta}_0), \delta + n),
$$

where

$$
W = (Z'A^{-1}Z + F)^{-1}Z'A^{-1}Y, \quad \hat{\boldsymbol{\beta}} = (Z'A^{-1}Z)^{-1}Z'A^{-1}Y, \quad \tilde{F} = Z'A^{-1}Z + F.
$$

Proof of Lemma 1. The proof follows arguments of Anderson (1984); see also Brown (1993).

Note. Using the identity

$$
(A + ZF^{-1}Z')^{-1} = A^{-1} - A^{-1}Z(F^{-1} + Z'A^{-1}Z)^{-1}Z'A^{-1},
$$

one can show that

$$(Y - Z\beta_0)'(A + ZF^{-1}Z')^{-1}(Y - Z\beta_0)$$

$$= (Y - Z\hat{\beta}_0)'A^{-1}(Y - Z\hat{\beta}_0) + (\hat{\beta} - \beta_0)'\left\{F^{-1} + (Z'A^{-1}Z)^{-1}\right\}^{-1}(\hat{\beta} - \beta_0),$$

reflecting the contributions from the likelihood and the prior distribution.

To state the next lemma, we let $Y = (Y^{[u]}, Y^{[g]})$, $Y^{[u]}$ and $Y^{[g]}$ having $n \times u$ and $n \times g$ dimensions, respectively. We adopt the following transformation of the partitioned covariance matrix Σ of Y:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \rightarrow (\Sigma_{22}, \tau, \Gamma)$$

for matrices $\Sigma_{11} : u \times u$, $\Sigma_{21} : g \times u$, $\Sigma_{22} : g \times g$ and $\tau = \Sigma_{22}^{-1}\Sigma_{21}$, $\Gamma = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

LEMMA 2. *Adopt the Gaussian and Generalized Inverted Wishart model specified by:*

$$\begin{cases} Y \mid \beta, \Sigma \sim N(Z\beta, A \otimes \Sigma), \\[2mm] \beta \mid \Sigma \sim N(\beta_0, F^{-1} \otimes \Sigma), \\[2mm] \tau \mid \Gamma \sim N(\tau_0, H \otimes \Gamma), \\[2mm] \Gamma \sim IW(\Psi_1, \delta_1), \\[2mm] \Sigma_{22} \sim IW(\Psi_2, \delta_2), \end{cases} \tag{13}$$

where $Z : n \times \ell$, $\beta = \left(\beta^{[u]}, \beta^{[g]}\right) : (\ell \times u, \ell \times g)$ and $\beta_0 = \left(\beta_0^{[u]}, \beta_0^{[g]}\right) : (\ell \times u, \ell \times g)$. Then the predictive distribution of $(Y^{[u]} \mid Y^{[g]})$ is

$$Y^{[u]} \mid Y^{[g]} \sim t_{n \times u}\left(\mu^{[u|g]}, \Phi^{[u|g]} \otimes \Psi^{[u|g]}, \delta_1 - u + 1\right)$$

where

$$\mu^{[u|g]} = Z\beta_0^{[u]} + \left(Y^{[g]} - Z\beta_0^{[g]}\right)\tau_0,$$

$$\Phi^{[u|g]} = A + ZF^{-1}Z' + \left(Y^{[g]} - Z\beta_0^{[g]}\right)H\left(Y^{[g]} - Z\beta_0^{[g]}\right)',$$

$$\Psi^{[u|g]} = (\delta_1 - u + 1)^{-1}\Psi_1.$$

Proof of Lemma 2. (i) Suppose $\beta = 0$. Then by standard results for the multivariate normal distribution, the conditional distribution of $(Y^{[u]} \mid Y^{[g]}, \Sigma)$, which does not depend on Σ_{22}, can be expressed as

$$\left(Y^{[u]} \mid Y^{[g]}, \tau, \Gamma\right) \sim N\left(Y^{[g]}\tau, I_n \otimes \Gamma\right).$$

Applying Lemma 1 to this distribution with the prior distributions of τ and Γ in (13) yields

$$\left(Y^{[u]} \mid Y^{[g]}\right) \sim t_{n \times u}\left(Y^{[g]}\tau_0, (\delta_1 - u + 1)^{-1}\left(I_n + Y^{[g]}HY^{[g]'}\right) \otimes \Psi_1, \delta_1 - u + 1\right).$$

(ii) Now suppose β follows the distribution in (13). Notice that

$$(A + ZF^{-1}Z')^{-1/2}(Y - Z\boldsymbol{\beta}_0) \mid \Sigma \sim N(0, I_n \otimes \Sigma).$$

The lemma follows immediately from the result in (i).

LEMMA 3. *In the setting of Lemma 2, assume further*

$$Y^{[g]} = \begin{pmatrix} Y^{[g]}_{(1)} \\ Y^{[g]}_{(2)} \end{pmatrix},$$

where the matrix $Y^{[g]}_{(1)} : m \times g$, $m < n$, holds the unobserved responses at the gauged sites. Let

$$\begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix} : \begin{pmatrix} m \times g \\ (n-m) \times g \end{pmatrix} = Z\boldsymbol{\beta}^{[g]}_0,$$

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} : \begin{pmatrix} m \times m & m \times (n-m) \\ (n-m) \times m & (n-m) \times (n-m) \end{pmatrix} = A + ZF^{-1}Z',$$

$$\mu_{(1|2)} = \mu_{(1)} + A_{12}(A_{22})^{-1}\left(Y^{[g]}_{(2)} - \mu_{(2)}\right),$$

$$\Phi_{(1|2)} = \frac{\delta_2 - g + 1}{\delta_2 - g + n - m + 1}\left\{A_{11} - A_{12}(A_{22})^{-1}A_{21}\right\},$$

$$\Psi_{(1|2)} = \frac{1}{\delta_2 - g + 1}\left\{\Psi_2 + \left(Y^{[g]}_{(2)} - \mu_{(2)}\right)'(A_{22})^{-1}\left(Y^{[g]}_{(2)} - \mu_{(2)}\right)\right\},$$

$$\delta_{(1|2)} = \delta_2 - g + n - m + 1.$$

Then the predictive distribution of $Y^{[g]}_{(1)}$ given data $Y^{[g]}_{(2)}$ is

$$\left(Y^{[g]}_{(1)} \mid Y^{[g]}_{(2)}, \mathcal{H}\right) \sim t_{m \times g}\left(\mu_{(1|2)}, \Phi_{(1|2)} \otimes \Psi_{(1|2)}, \delta_{(1|2)}\right). \tag{14}$$

Proof of Lemma 3. The GIW model (13) implies the GIW sub-model

$$\begin{cases} Y^{[g]} \mid \boldsymbol{\beta}^{[g]}, \Sigma_{22} \sim N(Z\boldsymbol{\beta}^{[g]}, A \otimes \Sigma_{22}), \\[2mm] \boldsymbol{\beta}^{[g]} \mid \Sigma_{22} \sim N(\boldsymbol{\beta}^{[g]}_0, F^{-1} \otimes \Sigma_{22}), \\[2mm] \Sigma_{22} \sim IW(\Psi_2, \delta_2). \end{cases}$$

Therefore, by Lemma 1

$$Y^{[g]} \sim t_{n \times g}\left[Z\boldsymbol{\beta}^{[g]}_0, (\delta_2 - g + 1)^{-1}(A + ZF^{-1}Z') \otimes \Psi_2, \delta_2 - g + 1\right]. \tag{15}$$

Conditioning the distribution in (15) upon $Y^{[g]}_{(2)}$ yields the predictive distribution (14).

Proof of Theorem 1. Part (i) of the proof is a straightforward application of Lemma 3. For part (ii), the distribution is obtained by first applying Lemma 2 to $Y^{[g_j]}$ conditional on $Y^{[g_{j+1},\ldots,g_k]}$ and then applying Lemma 3 to $Y^{[g_j^m]}$ conditional on $Y^{[g_j^o]}$ and $Y^{[g_{j+1},\ldots,g_k]}$. Part (iii) is an immediate result of Lemma 2.

Proof of Theorem 2. (i) Model (1) implies

$$
\begin{cases}
\beta^{[g_k]} \mid \Sigma_{kk}, \beta_0, F \sim N\left(\beta_0^{[k]}, F^{-1} \otimes \Sigma_{kk}\right), \\
Y^{[g_k^o]} \mid \beta^{[g_k]}, \Sigma_{kk}, \mathcal{H} \sim N\left(Z_{(k)}\beta^{[g_k]}, I_{n-m_k} \otimes \Sigma_{kk}\right),
\end{cases}
\tag{16}
$$

which gives the posterior distribution of $(\beta^{[g_k]} \mid D, \Sigma_{kk}, \mathcal{H})$, as in (8), by means of Lemma 1.
  Similarly, model (1) implies

$$
\begin{cases}
\beta^{[g_j]} \mid \beta^{[g_{j+1},\ldots,g_k]}, \tau_j, \Gamma_j, \mathcal{H} \sim N\left(\beta_0^{[g_j]} + \left(\beta^{[g_{j+1},\ldots,g_k]} - \beta_0^{[g_{j+1},\ldots,g_k]}\right)\tau_j, F^{-1} \otimes \Gamma_j\right), \\
Y^{[g_j^o]} \mid Y_{(j)}^{[g_{j+1},\ldots,g_k]}, \beta^{[g_j,g_{j+1},\ldots,g_k]}, \tau_j, \Gamma_j, \mathcal{H} \sim N\left(Z_{(j)}\beta^{[g_j]} + \tilde{\varepsilon}_{(j)}^{[g_{j+1},\ldots,g_k]}\tau_j, F^{-1} \otimes \Gamma_j\right),
\end{cases}
\tag{17}
$$

where

$$
\tilde{\varepsilon}_{(j)}^{[g_{j+1},\ldots,g_k]} = Y_{(j)}^{[g_{j+1},\ldots,g_k]} - Z_{(j)}\beta_0^{[g_{j+1},\ldots,g_k]}.
$$

Applying Lemma 1 again gives the posterior of $(\beta^{[g_j]} \mid D, \beta^{[g_{j+1},\ldots,g_k]}, \tau_j, \Gamma_j, \mathcal{H})$ as in (8).
  (ii) Combining (16), (17) and (2) and using the result of Lemma 1 yields distribution (9).

Proof of Corollary 1. Taking conditional expectations of the β's given $(D, \mathcal{H})$ in Theorem 2 yields the result.

APPENDIX B: CALCULATION OF POSTERIOR EXPECTATIONS

LEMMA 4 (This is the same as Corollary 1 in the text). *The posterior means of β given data D and hyperparameter set $\mathcal{H}$ are given below:*

$$
\mathrm{E}\left(\beta^{[g_k]} \mid D, \mathcal{H}\right) = \tilde{\beta}^{[g_k^o]},
$$

$$
\mathrm{E}\left(\beta^{[g_j]} \mid D, \mathcal{H}\right) = \tilde{\beta}^{[g_j^o]} + \left\{\mathrm{E}\left(\beta^{[g_{j+1},\ldots,g_k]} \mid D, \mathcal{H}\right) - \tilde{\beta}^{[g_{j+1},\ldots,g_k]}\right\}\tilde{\tau}_j,
$$

where

$$
\mathrm{E}\left(\beta^{[g_{j+1},\ldots,g_k]} \mid D, \mathcal{H}\right) = \left(\mathrm{E}\left(\beta^{[g_{j+1}]} \mid D, \mathcal{H}\right), \ldots, \mathrm{E}\left(\beta^{[g_k]} \mid D, \mathcal{H}\right)\right)
$$

is computed in a recursive fashion, for $j = 1, \ldots, k-1$.

Proof of Lemma 4. Straightforward using Theorem 2 and a conditional expectation argument.

LEMMA 5. *The posterior mean of $\Sigma^{-1} = \Sigma_{(11)}^{-1}$ given data D and hyperparameter set $\mathcal{H}$ is computed recursively as follows:*

$$
\mathrm{E}\left(\Sigma_{(jj)}^{-1} \mid D, \mathcal{H}\right) = \begin{pmatrix} \tilde{\delta}_j \tilde{\Psi}_j^{-1} & -\tilde{\delta}_j \tilde{\Psi}_j^{-1}\tilde{\tau}_j' \\ -\tilde{\delta}_j \tilde{\tau}_j \tilde{\Psi}_j^{-1} & \tilde{\delta}_j \tilde{\tau}_j \tilde{\Psi}_j^{-1}\tilde{\tau}_j' + g_j \tilde{H}_j + \mathrm{E}\left(\Sigma_{(j+1,j+1)}^{-1} \mid D, \mathcal{H}\right) \end{pmatrix},
$$

for $j = 1, \ldots, k-1$, and $\mathrm{E}\left(\Sigma_{(kk)}^{-1} \mid D, \mathcal{H}\right) = \tilde{\delta}_k \tilde{\Psi}_k^{-1}$.

Proof of Lemma 5. Notice that

$$\Sigma^{-1}_{(jj)} = \left(\Gamma_j^{-1} \quad -\Gamma_j^{-1}\tau_j' - \tau_j \Gamma_j^{-1} \quad \tau_j \Gamma_j^{-1}\tau_j' + \Sigma^{-1}_{(j+1,j+1)} \right) \text{ for } j = 1, \ldots, k-1.$$

Using Theorem 2,

$$
\begin{aligned}
\mathrm{E}\left(\Gamma_j^{-1} \mid D, \mathcal{H} \right) &= \tilde{\delta}_j \tilde{\Psi}_j^{-1}, \\
\mathrm{E}\left(\Gamma_j^{-1}\tau_j' \mid D, \mathcal{H} \right) &= \tilde{\delta}_j \tilde{\Psi}_j^{-1}\tilde{\tau}_j', \\
\mathrm{E}\left(\tau_j \Gamma_j^{-1}\tau_j' \mid D, \mathcal{H} \right) &= \tilde{\delta}_j \tilde{\tau}_j \tilde{\Psi}_j^{-1}\tilde{\tau}_j' + g_j \tilde{H}_j,
\end{aligned}
$$

for $j = 1, \ldots, k-1$, and $\mathrm{E}\left(\Sigma^{-1}_{(kk)} \mid D, \mathcal{H} \right) = \mathrm{E}\left(\Sigma^{-1}_{kk} \mid D, \mathcal{H} \right) = \tilde{\delta}_k \tilde{\Psi}_k^{-1}$. This proves the lemma.

LEMMA 6. *Let ψ denote the digamma function. Then we have*

$$\mathrm{E}\left(\log |\Sigma_{kk}| \mid D, \mathcal{H} \right) = -g_k \log 2 - \sum_{i=1}^{g_k} \psi\left(\frac{\tilde{\delta}_k - i + 1}{2} \right) + \log |\tilde{\Psi}_k|,$$

and

$$\mathrm{E}\left(\log |\Gamma_j| \mid D, \mathcal{H} \right) = -g_j \log 2 - \sum_{i=1}^{g_j} \psi\left(\frac{\tilde{\delta}_j - i + 1}{2} \right) + \log |\tilde{\Psi}_j|.$$

Proof of Lemma 6. See Chen (1979).

LEMMA 7. *The posterior expectation $\mathrm{E}\left(\beta\Sigma^{-1} \mid D, \mathcal{H} \right)$ is obtained as follows:*
$$\mathrm{E}\left(\beta\Sigma^{-1} \mid D, \mathcal{H} \right) = \mathrm{E}\left(\beta^{[g_1,\ldots,g_k]}\Sigma^{-1}_{(11)} \mid D, \mathcal{H} \right),$$
$$\mathrm{E}\left(\beta^{[g_j,\ldots,g_k]}\Sigma^{-1}_{(jj)} \mid D, \mathcal{H} \right) = \left(\tilde{\delta}_j \left(\tilde{\beta}^{[g_j^o]} - \tilde{\beta}^{[g_{j+1},\ldots,g_k]}\tilde{\tau}_j \right) \tilde{\Psi}_j^{-1}, \right.$$
$$\mathrm{E}\left(\beta^{[g_{j+1},\ldots,g_k]}\Sigma^{-1}_{(j+1,j+1)} \mid D, \mathcal{H} \right) - \tilde{\delta}_j \left(\tilde{\beta}^{[g_j^o]} - \tilde{\beta}^{[g_{j+1},\ldots,g_k]}\tilde{\tau}_j \right) \tilde{\Psi}_j^{-1}\tilde{\tau}_j' + g_j \tilde{\beta}^{[g_{j+1},\ldots,g_k]}\tilde{H}_j \right),$$
and finally $\mathrm{E}\left(\beta^{[g_k]}\Sigma^{-1}_{kk} \mid D, \mathcal{H} \right) = \tilde{\delta}_k \tilde{\beta}^{[g_k]}\tilde{\Psi}_k^{-1}.$

Proof of Lemma 7. We can write

$$\beta^{[g_j,\ldots,g_k]}\Sigma^{-1}_{(jj)}$$
$$= \left(\left(\beta^{[g_j]} - \beta^{[g_{j+1},\ldots,g_k]}\tau_j \right) \Gamma_j^{-1}, \beta^{[g_{j+1},\ldots,g_k]}\Sigma^{-1}_{(j+1,j+1)} - \left(\beta^{[g_j]} - \beta^{[g_{j+1},\ldots,g_k]}\tau_j \right) \Gamma_j^{-1}\tau_j' \right).$$

Taking expectations and using this result recursively yield the lemma.

LEMMA 8. *The posterior expectation $\beta\Sigma^{-1}\beta'$ is as follows:*

$$
\begin{aligned}
\mathrm{E}\left(\beta\Sigma^{-1}\beta' \mid D, \mathcal{H} \right) =\ & \tilde{\delta}_k \tilde{\beta}^{[g_k^o]}\tilde{\Psi}_k^{-1}\tilde{\beta}^{[g_k^o]'} \\
& + \sum_{j=1}^{k-1} \tilde{\delta}_j \left(\tilde{\beta}^{[g_j^o]} - \tilde{\beta}^{[g_{j+1},\ldots,g_k]}\tilde{\tau}_j \right) \tilde{\Psi}_j^{-1} \left(\tilde{\beta}^{[g_j^o]} - \tilde{\beta}^{[g_{j+1},\ldots,g_k]}\tilde{\tau}_j \right)' \\
& + \sum_{j=1}^{k-1} g_j \tilde{\beta}^{[g_{j+1},\ldots,g_k]}\tilde{H}_j \tilde{\beta}^{[g_{j+1},\ldots,g_k]'} + \sum_{j=1}^{k} g_j \tilde{F}_j^{-1}.
\end{aligned}
$$

Proof of Lemma 8. The proof is obtained by observing that

$$\beta\Sigma^{-1}\beta' = \beta^{[g_k]}\Sigma_{kk}^{-1}\beta^{[g_k]'} + \sum_{j=1}^{k-1} \left(\beta^{[g_j]} - \beta^{[g_{j+1},\dots,g_k]}\tau_j\right)\Gamma_j^{-1}\left(\beta^{[g_j]} - \beta^{[g_{j+1},\dots,g_k]}\tau_j\right)'.$$

Taking expectation given $(D, \mathcal{H})$ on both sides of above leads to the result.

ACKNOWLEDGEMENTS

REFERENCES

T. W. Anderson (1984). *An Introduction to Multivariate Statistical Analysis*, Second Edition. Wiley, New York.

M. S. Bartlett (1933). On the theory of statistical regression. *Proceedings of the Royal Society of Edinburgh*, 53, 260–283.

P. J. Brown (1993). *Measurement, Regression, and Calibration*. Clarendon, Oxford.

P. J. Brown, N. D. Le & J. V. Zidek (1994a). Multivariate spatial interpolation and exposure to air pollutants. *The Canadian Journal of Statistics*, 22, 489–510.

P. J. Brown, N. D. Le & J. V. Zidek (1994b). Inference for a covariance matrix. In *Aspects of Uncertainty: A Tribute to D. V. Lindley* (P. R. Freeman & A. F. M. Smith, eds.), Wiley, Chichester, pp. 77–92.

R. J. Carroll, D. Ruppert & D. Stefanski (1995). *Measurement Error in Nonlinear Models*. Chapman & Hall, London.

C. F. Chen (1979). Bayesian Inference for a normal dispersion matrix and its application to stochastic multiple regression analysis. *Journal of the Royal Statistical Society Series B*, 41, 235–248.

N. C. Cressie (1991). *Statistics for Spatial Data*. Wiley, New York.

V. De Oliveira, K. Benjamin & D. A. Short (1997). Bayesian prediction of transformed Gaussian random fields. *Journal of the American Statistical Association*, 92, 1422–1433.

A. P. Dempster, N. M. Laird & D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society Series B*, 39, 1–38.

C. Duddek, N. D. Le, J. V. Zidek & R. T. Burnett (1995). Multivariate imputation in cross sectional analysis of health effects associated with air pollution (with discussion). *Journal of Environmental and Ecological Statistics*, 2, 191–212.

M. Gaudard, M. Karson, E. Linder & D. Sinha (1999). Bayesian spatial prediction (with discussion). *Journal of Ecological and Environmental Statistics*, 6, 147–171.

T. C. Haas (1996). Multivariate spatial prediction in the presence of nonlinear trend and covariance nonstationary. *Environmetrics*, 7, 145–165.

M. S. Handcock & M. L. Stein (1993). A Bayesian analysis of kriging. *Technometrics*, 35, 403–410.

M. S. Handcock & J. Wallis (1994). An approach to statistical spatial-temporal modeling of meteorological fields (with discussion and rejoinder). *Journal of the American Statistical Association*, 89, 368–390.

N. D. Le, L. Sun & J. V. Zidek (1998). A note on the existence of maximum likelihood estimates for Gaussian-inverted Wishart models. *Statistics & Probability Letters*, 40, 133–137.

N. D. Le, W. Sun & J. V. Zidek (1997). Bayesian multivariate spatial interpolation with data missing-by-design. *Journal of the Royal Statistical Society Series B*, 59, 501–510.

N. D. Le & J. V. Zidek (1992). Interpolation with uncertain spatial covariances: a Bayesian alternative to kriging. *Journal of Multivariate Analysis*, 43, 351–74.

N. D. Le & J. V. Zidek (1994). Network designs for monitoring multivariate random spatial fields. In *Recent Advances in Statistics and Probability: Proceedings of the 4th International Meeting of Statistics in the Basque Country, San Sebastián, Spain, 4–7 August 1992* (J. P. Vilaplana & M. L. Puri, eds.), VSP, Utrecht, pp. 191–206.

K. H. Li, N. D. Le, L. Sun & J. V. Zidek (1999). Spatial-temporal models for ambient hourly PM_{10} in Vancouver. *Environmetrics*, 10, 321–338.

R. J. A. Little & D. B. Rubin (1987). *Statistical Analysis with Missing Data*. Wiley, New York.

C. Liu (1996). Bayesian robust multivariate linear regression with incomplete data. *Journal of the American Statistical Association*, 91, 1219–1227.

C. Liu (1999). Efficient ML estimation of the multivariate normal distribution from incomplete data. *Journal of Multivariate Analysis*, 69, 206–217.

D. A. Pierce, D. O. Stram, M. Vaeth & D. W. Schafer (1992). The errors in variables problem: considerations provided by radiation dose-response analyses of A-bomb survivor data. *Journal of the American Statistical Association*, 87, 351–359.

D. B. Rubin & J. L. Shaffer (1990). Efficiently creating multiple imputations for incomplete multivariate normal data. In *Proceedings of the Statistical Computing Section*, American Statistical Association, pp. 83–88.

P. D. Sampson & P. Guttorp (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87, 108–119.

W. Sun (1998). Comparison of a co-kriging method with a Bayesian alternative. *Environmetrics*, 9, 445–457.

W. Sun, N. D. Le, J. V. Zidek & R. Burnett (1998). Assessment of a Bayesian multivariate interpolation approach for health impact studies. *Environmetrics*, 9, 565–586.

J. V. Zidek (1997). Interpolating air pollution for health impact assessment. In *Statistics for the Environment 3: Pollution Assessment and Control* (V. Barnett & K. F. Turkman, eds.), Wiley, Chichester, pp. 251–268.

J. V. Zidek, W. Sun & N. D. Le (2000). Designing and integrating composite networks for monitoring multivariate Gaussian pollution fields. *Journal of the Royal Statistical Society Series C*, 49, 63–79.

J. V. Zidek, R. White, N. D. Le, W. Sun & R. T. Burnett (1998). Imputing unmeasured explanatory variables in environmental epidemiology with application to health impact analysis of air pollution. *Journal of Environmental and Ecological Statistics*, 5, 99–115.

Nhu LE: nhu@stat.ubc.ca

Cancer Control Research

British Columbia Cancer Agency, 600 West 10th Avenue

Vancouver, British Columbia, Canada V5Z 4E6

Li SUN: lisun@stat.ubc.ca

James V. ZIDEK: jim@stat.ubc.ca

Department of Statistics, University of British Columbia

Vancouver, British Columbia, Canada V6T 1Z2