Bayesian Multivariate Spatial Interpolation with Data Missing by Design

Author(s): Nhu D. Le, Weimin Sun and James V. Zidek

Source: *Journal of the Royal Statistical Society. Series B (Methodological)*, 1997, Vol. 59, No. 2 (1997), pp. 501-510

Published by: Wiley for the Royal Statistical Society

Stable URL: https://www.jstor.org/stable/2346059

*J. R. Statist. Soc.* B (1997)
**59**, *No.* 2, *pp.* 501–510

# Bayesian Multivariate Spatial Interpolation with Data Missing by Design

By NHU D. LE, WEIMIN SUN and JAMES V. ZIDEK†

*University of British Columbia, Vancouver, and BC Cancer Agency, Vancouver, Canada*

*Statistics Canada, Ottawa, Canada*

*University of British Columbia, Vancouver, Canada*

[Received September 1994. Final revision August 1996]

SUMMARY

In a network of $s_g$ sites, responses like levels of airborne pollutant concentrations may be monitored over time. The sites need not all measure the same set of response items and unmeasured items are considered as data *missing by design*. We propose a hierarchical Bayesian approach to interpolate the levels of, say, $k$ responses at $s_u$ other locations called ungauged sites and also the unmeasured levels of the $k$ responses at the gauged sites. Our method involves two steps. First, when all hyperparameters are assumed to be known, a predictive distribution is derived. In turn, an interpolator, its variance and a simultaneous interpolation region are obtained. In step two, we propose the use of an empirical Bayesian approach to estimate the hyperparameters through an EM algorithm. We base our theory on a linear Gaussian model and the relationship between a multivariate normal and matrix $T$-distribution. Our theory allows us to pool data from several existing networks that measure different subsets of response items for interpolation.

*Keywords*: BAYESIAN INTERPOLATION; CO-KRIGING; MATRIX $T$-DISTRIBUTION; PREDICTIVE DISTRIBUTION; SPATIAL INTERPOLATION

## 1. INTRODUCTION

Environmental studies lead naturally to the statistical analysis of random space–time response fields. These analyses rely on measurements made at ambient monitors in a spatial network or combination of networks. Practical considerations can result in the monitors or 'gauged sites' being widely separated. Yet values of the response fields between these sites (i.e. at the 'ungauged sites') may be needed along with their spatial predictive distribution to account for the uncertainty involved in their interpolation.

Spatial interpolation problems arise in diverse fields such as mining, engineering, geology, soil science and hydrology. Analysts commonly tackle such problems with the well-known method of *kriging*, introduced in the 1960s by Matheron (Cressie, 1991). Kriging relies on the variogram, i.e. the mean-squared difference of the random responses at two locations, which is assumed to depend on the location of the sites only through the distance separating them (i.e. to be isotropic). Kriging has an appealing optimality property when the variogram is known, that of being the best linear unbiased estimator.

Improvement in the accuracy of the geostatistical method of kriging is possible

†*Address for correspondence*: Department of Statistics, University of British Columbia, 6356 Agriculture Road, Vancouver, British Columbia, V6T 1Z2, Canada.
E-mail: jim@stat.ubc.ca

 0035–9246/97/59501

when additional information from covariables can be incorporated in the interpolator. Then the method is called *co-kriging* (see Haas (1990)).

However, kriging and co-kriging lose their optimality when the variogram must be estimated. Moreover, not all available data can be used in doing so since these geostatistical methods are intended for use at a specified time point. In fact, unrealistically simple isotropic variogram models are used so that they can be fitted with cross-sectional data alone.

This paper presents an extension of the work of Le and Zidek (1992) and Brown *et al.* (1994a) to gain an interpolation theory for multivariate random spatial fields when the gauged sites have data *missing by design*. Like its predecessors that extension incorporates uncertainty about model parameters including those in the spatial covariance matrix. Although we are forced to make empirical compromises in our Bayesian approach, our method unlike the kriging approaches allows fully arbitrary spatial covariance structures.

Our approach yields not just an interpolator but a multivariate $t$-posterior spatial predictive distribution for the responses at the ungauged sites and the missing by design responses at the gauged sites. That distribution can be dynamically updated with the incoming longitudinal data. Moreover it can be used in conjunction with non-linear regression analyses like that described at the beginning of this section to deal with structural errors in variables.

We state our main results in Section 2 and a procedure for parameter estimation in Section 3. We discuss our findings in Section 4. All proofs and detailed derivations have been omitted. Most of the details can be found in Sun (1994).

## 2. INTERPOLATION WITH DATA MISSING BY DESIGN

Suppose that $s_g$ gauged sites have been established within a geographical area. With each such site $i = 1, \ldots, s_g$ we associate a response vector of items, a fixed subset of which is measured at every observation time $t = 1, \ldots, n$.

We take our objective as that of imputing responses or item measurements missing by design including those at $s_u$ hypothetical ungauged sites. We are to do this using the measurements which have been made, assuming that the measurement series are complete over time. In all we have $s = s_g + s_u$ sites in the region at which some (not necessarily proper) subset of item measurements will be missing by design, that subset being the same for all $t$.

Our Bayesian approach to interpolation involves two steps. In the first step, described in the remainder of this section, we derive the required Bayesian interpolator assuming that the hyperparameters are known. In the second step, described in the next section, we fit the hyperparameters involved. To simplify our presentation, we shall rely on the notation of Dawid (1981) as presented by Brown (1993) with requisite theory. Thus, for a random matrix $U$ $(p \times q)$, $U \sim \mu + \mathcal{N}(A, B)$ would mean that $U$ has a vector normal distribution with matrix mean $\mu$; each row of $U$ would have covariance matrix $A$; each column, covariance matrix $B$. Furthermore, for a random matrix $V$ $(q \times q)$, $V \sim \mathcal{IW}(m^*; Q)$ means that $V^{-1}$ has a Wishart distribution with $m^* + q - 1$ degrees of freedom and scale matrix $Q^{-1}$. Finally, for a random matrix $T$ $(p \times q)$, $T \sim \mathcal{T}(m^*; P, Q)$ means that $T|V \sim \mathcal{N}(P, V)$ and $V \sim \mathcal{IW}(m^*; Q)$.

For each time point $t = 1, \ldots, n$, let $X_t$ represent the $(1 \times sk)$-dimensional random

vector of all response co-ordinates at both gauged and ungauged sites. Like Le and Zidek (1992) and Brown *et al.* (1994a), we assume a normal model for the conditional sampling distribution. To be precise, letting $X' = (X_1', \ldots, X_n')$, we assume

$$X \sim ZB + \mathcal{N}(I_n, \Sigma). \tag{1}$$

The covariate matrix $Z$ $(n \times h)$ determines the mean of $X$ through the $h \times sk$ matrix of regression coefficients $B$.

We use the (conjugate) priors for $B$ and $\Sigma$ adopted in Le and Zidek (1992) and Brown *et al.* (1994a):

$$B \sim B^0 + \mathcal{N}(F^{-1}, \Sigma) \tag{2}$$

and

$$\Sigma \sim IW(\delta, \Phi). \tag{3}$$

Partition $X_t$ as $X_t = (X_t^{(u)}, X_t^{(g)})$, $X_t^{(u)}$ $(1 \times s_u k)$ being the concatenation of the response vectors of all the ungauged sites at time $t$ and $X_t^{(g)}$ $(1 \times s_g k)$ that of the gauged sites. After appropriate rearrangement of its elements, $X_t^{(g)}$ can further be partitioned into $X_t^{(1)}$ $(1 \times l)$ and $X_t^{(2)}$ $(1 \times (s_g k - l))$, corresponding respectively to the items unmeasured and measured by design at the gauged sites. Since the same $l$ items remain unmeasured during the whole monitoring period and they comprise $l$ fixed columns in the matrix $X$, we call them *missing columns* in what follows for expository simplicity.

Represent the unmeasured and measured column numbers of $X^{(g)}$ by $i_1, \ldots, i_l$ and $i_{l+1}, \ldots, i_{s_g k}$ respectively. Let $r_j$, $j = 1, \ldots, s_g k$ be an $(s_g k \times 1)$-dimensional vector with $j$th element 1 and the rest 0. We can then construct indicator matrices $R_1$ and $R_2$, which 'mark' the position of missing and present columns respectively. More precisely, $R_1 = (r_{i_1}, \ldots, r_{i_l})$ and $R_2 = (r_{i_{l+1}}, \ldots, r_{i_{s_g k}})$. Finally $R$ denotes the orthogonal matrix $R = (R_1, R_2)$. Observe that $X^{(1)} = X^{(g)} R_1$ and $X^{(2)} = X^{(g)} R_2$ consist of just the missing and present columns respectively for the gauged sites. Because the response vector $X_t$ has been partitioned into three parts, we partition $B$, $\Sigma$, $B^0$ and $\Phi$ accordingly. For example, we first partition $\Sigma$ as

$$\Sigma = \begin{pmatrix} \Sigma_{uu} & \Sigma_{ug} \\ \Sigma_{gu} & \Sigma_{gg} \end{pmatrix},$$

where $\Sigma_{uu}$ and $\Sigma_{gg}$ are $s_u k \times s_u k$ and $s_g k \times s_g k$ matrices respectively. $R' \Sigma_{gg} R$ can be further partitioned as

$$R' \Sigma_{gg} R = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} R_1' \Sigma_{gg} R_1 & R_1' \Sigma_{gg} R_2 \\ R_2' \Sigma_{gg} R_1 & R_2' \Sigma_{gg} R_2 \end{pmatrix},$$

where $\Sigma_{11}$ and $\Sigma_{22}$ are $l \times l$ and $(s_g k - l) \times (s_g k - l)$ matrices respectively. Further, let

$$\Psi_{gg} = R' \Phi_{gg} R = \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix} = \begin{pmatrix} R_1' \Phi_{gg} R_1 & R_1' \Phi_{gg} R_2 \\ R_2' \Phi_{gg} R_1 & R_2' \Phi_{gg} R_2 \end{pmatrix},$$

with $\Psi_{11}$ being $l \times l$ and $\Psi_{22}$ being $(s_g k - l) \times (s_g k - l)$. Let $\Psi_{1|2} = \Psi_{11} - \Psi_{12} \Psi_{22}^{-1} \Psi_{21}$ and $\gamma = \Psi_{22}^{-1} \Psi_{21}$ for use in what follows. Finally, let

$$(B_1^0,\ B_2^0) = B_g^0 R = (B_g^0 R_1,\ B_g^0 R_2).$$

We obtain our Bayesian interpolators from predictive distributions. Conditionally on $X^{(2)} = x^{(2)}$, the theory given by Brown (1993), appendix A.5, yields the predictive distribution of the levels of the unmeasured items at the gauged sites:

$$X^{(1)} \sim ZB_1^0 + (x^{(2)} - ZB_2^0)\gamma + \mathcal{T}(\delta + s_g k - l,\ P_{1|2},\ \Psi_{1|2})$$

where

$$P_{1|2} = I_n + ZF^{-1}Z' + (x^{(2)} - ZB_2^0)\Psi_{22}^{-1}(x^{(2)} - ZB_2^0)'$$

and

$$\Psi_{1|2} = \Psi_{11} - \Psi_{12}\Psi_{22}^{-1}\Psi_{21}.$$

Interpolating at the ungauged sites conditionally on $X^{(2)} = x^{(2)}$ requires another result which follows immediately from general theory:

$$X^{(u)} \sim ZB_u^0 + (x^{(2)} - ZB_2^0)\Psi_{22}^{-1}R_2'\Phi_{gu} + \mathcal{T}(\delta + s_g k - l,\ P_{u|2},\ \Phi_{u|2})$$

where

$$P_{u|2} = I_n + ZF^{-1}Z' + (x^{(2)} - ZB_2^0)\Psi_{22}^{-1}(x^{(2)} - ZB_2^0)'$$

and

$$\Phi_{u|2} = \Phi_{uu} - \Phi_{ug}R_2\Psi_{22}^{-1}R_2'\Phi_{gu}.$$

The joint conditional predictive distribution of $(X^{(u)},\ X^{(1)})|X^{(2)} = x^{(2)}$ can be derived in the same way, but we omit the details for brevity.

By applying these results, we find $X_t^{(u)}|X^{(2)} = x^{(2)}$ to be a multivariate $t$-distribution, a special case of a matrix $T$-distribution:

$$X_t^{(u)} \sim Z_t B_u^0 + (x_t^{(2)} - Z_t B_2^0)\Psi_{22}^{-1}R_2'\Phi_{gu} + \mathcal{T}\big(\delta + s_g k - l,\ P_{t|2},\ \Phi_{u|2}\big), \tag{4}$$

where

$$P_{t|2} = 1 + Z_t F^{-1}Z_t' + (x_t^{(2)} - Z_t B_2^0)\Psi_{22}^{-1}(x_t^{(2)} - Z_t B_2^0)'$$

and $\Phi_{u|2}$ is defined above. A similar result can be obtained for $X_t^{(1)}|X^{(2)} = x^{(2)}$. We can now find the required Bayesian interpolators from these conditional predictive distributions:

$$E(X_t^{(1)}|X^{(2)} = x^{(2)}) = Z_t B_1^0 + (x_t^{(2)} - Z_t B_2^0)\gamma, \tag{5}$$

$$E(X_t^{(u)}|X^{(2)} = x^{(2)}) = Z_t B_u^0 + (x_t^{(2)} - Z_t B_2^0)\Psi_{22}^{-1}R_2'\Phi_{gu}. \tag{6}$$

We would expect that, as Brown (1993), p. 133, implicitly suggests, interest will often focus on the specific case of $t = n + 1$, representing the future. However, in some applications like that described at the beginning of Section 1, interest obtains in all the missing values. Then the entire unmeasured field must be reconstructed while reflecting the uncertainties which obtain from that reconstruction.

We could simplify our task and attain the special case of $s_u k = 1$ by interpolating

values item by item and site by site for all items and all sites. The multivariate $t$-distribution in expression (4) then reduces to the univariate $t$-distribution. A direct application of a variance formula for the multivariate $t$-distribution (Press, 1982) to the distribution $X_t^{(u)}|X^{(2)} = x^{(2)}$ leads to a formula for the variance of our Bayesian interpolator:

$$\operatorname{var}(X_t^{(u)}|X^{(2)} = x^{(2)}) = (\delta + s_g k - l - 2)^{-1} P_{t|2} \Phi_{u|2}. \tag{7}$$

Note, as Brown (1993) does, the missing scale factor, $\delta + s_g k - l$ in equation (7), the result of a disagreement between the classical definitions of the $t$-distribution and that of Dawid (1981) in the univariate case. Note also that this variance decreases as $s_g k - l$ increases, as might be expected on heuristic grounds.

With this variance, we could derive a pointwise posterior credibility interval for the interpoland. We could then move on to address another item or site. Continuing sequentially in this fashion, we would eventually interpolate the entire network of ungauged and gauged sites and supply say 95% pointwise credibility intervals for all sites and items in the process. Indeed this process would mimic what has been done in conventional kriging analysis.

However, the univariate approach ignores the multiplicity of the imputations which have been simultaneously carried out over this correlated spatial field. In particular, that approach ignores the effect on the nominal coverage probability of those pointwise intervals.

We have available a more satisfactory approach in as much as the posterior distribution of $X_t^{(u)}|X^{(2)} = x^{(2)}$ in expression (4) allows us to derive a simultaneous credibility region. To state our result for the ungauged sites, we let $\hat{x}_t^{(u)}$ denote the Bayesian interpolator of $X_t^{(u)}$ when there are data missing by design. More precisely, $\hat{x}_t^{(u)} = Z_t B_u^0 + (x_t^{(2)} - Z_t B_2^0)\Psi_{22}^{-1} R_2' \Phi_{gu}$. Then, given $X^{(2)} = x^{(2)}$, the $(1 - \alpha)$-level $(0 < \alpha < 1)$ simultaneous posterior credibility region is $\{X_t^{(u)}: (X_t^{(u)} - \hat{x}_t^{(u)})\Phi_{u|2}^{-1}(X_t^{(u)} - \hat{x}_t^{(u)})' < b\}$ where,

$$b = (s_u k * P_{t|2} * F_{1-\alpha, s_u k, \delta-(s_u - s_g)k-l}) * \{\delta - (s_u - s_g)k - l\}^{-1}.$$

Notice the price paid in degrees of freedom as $s_u$ and $k$ increase. That price increases particularly rapidly since these numbers enter through their product and soon use all the prior information expressed through $\delta$ and that gained from the gauged sites expressed through $s_g k - l$.

## 3. ESTIMATION OF HYPERPARAMETERS

To carry out multivariate interpolation, the hyperparameters $\Phi$, $\delta$, $F$ and $B^0$ must be specified. In a completely Bayesian hierarchical approach another level of priors would be needed to express our uncertainty about those hyperparameters. An alternative approach, which is suitable when the parameters are not sensitive to their prior specification, entails the use of type II maximum likelihood estimation.

### 3.1. *Estimation of* $\Phi$ *and* $\delta$

Like Brown *et al.* (1994a) we may adopt the EM algorithm to find the type II maximum likelihood estimate of the hyperparameters in $\Phi$ and $\delta$. Furthermore, to

reduce the number of parameters estimated, we, like Brown *et al.* (1994a), adopt a Kronecker structure, $\Phi = \Lambda \otimes \Omega$, $\Lambda$ corresponding to covariation between sites and $\Omega$ between response items. The estimation of $\Phi$ takes two steps:

(a) $\Lambda_g$, $\Omega$ and $\delta$ are estimated by the EM algorithm where $\Lambda_g$ denotes the submatrix of $\Lambda$ corresponding to the gauged sites;
(b) a nonparametric approach of Sampson and Guttorp (1992) is used to extend $\Lambda_g$ to $\Lambda$.

When data are missing by design we cannot apply the EM method of Brown *et al.* (1994a) directly. A modification is needed. Here we sketch only the modified part of the procedure for estimating $\Phi$ and $\delta$. Refer to Brown *et al.* (1994a) and Sun (1994) for a detailed description of the procedure as a whole.

To state the modification, we start with the original EM algorithm of Dempster *et al.* (1977). When the sampling distribution $X|\Phi$ is from a regular exponential family, the $p$th iteration of the EM algorithm consists of

(a) the *E-step*—estimate the complete data sufficient statistics $t(x)$ by finding data $x$ and $t^{(p)} = E\{t(X)|y, \Delta\}$, where $y$ is the observed part of the data $x$—and
(b) the *M-step*—determine $\Delta^{(p+1)}$ as the solution of the equation $E\{t(X)|\Delta\} = t^{(p)}$.

Wu (1983) proved that the point of convergence $\Delta$ does maximize the likelihood function.

In our case, the complete data set is $x = \{x^{(1)}, x^{(2)}, \Sigma_{gg}\}$, the observed data set $y = x^{(2)}$ and the parameter set $\Delta = \{\Phi, \delta\}$, and under model (1)–(3) $\{\Sigma_{gg}^{-1}, \log|\Sigma_{gg}|\}$ is the sufficient statistic for $\Phi$ and $\delta$ (see Chen (1979)). So given the current values of $\Phi$ and $\delta$ at the E-step we need to find

$$E(\Sigma_{gg}^{-1}|X^{(2)} = x^{(2)}, \text{ and } \Phi_{gg}, \delta),$$

$$E(\log|\Sigma_{gg}|\,|X^{(2)} = x^{(2)}, \Phi_{gg}, \delta)$$

where $\Phi_{gg} = \Lambda_g \otimes \Omega$.

Noting that $\Psi_{ij} = R_i' \Phi_{gg} R_j$, we find

$$E(\Sigma_{gg}^{-1}|X^{(2)} = x^{(2)}, \Phi_{gg}, \delta) = R\begin{pmatrix} \{\delta + (s_g - s_u)k - 1\}\Psi_{1|2}^{-1} & -\{\delta + (s_g - s_u)k - 1\}\Psi_{1|2}^{-1}\gamma' \\ -\{\delta + (s_g - s_u)k - 1\}\gamma\Psi_{1|2}^{-1} & d_1 \end{pmatrix} R'$$

where

$$G_1 = I - Z(Z'Z)^{-1}Z',$$

$$G_2 = X^{(2)\prime}Z(Z'Z)^{-1} - B_2^0,$$

$$\tilde{S} = X^{(2)\prime}G_1X^{(2)} + G_2'\{(Z'Z)^{-1} + F^{-1}\}^{-1}G_2,$$

$$\hat{\Psi}_{22} = \Psi_{22} + \tilde{S},$$

$$d_1 = \{\delta + (s_g - s_u)k + n - l - h - 1\}\hat{\Psi}_{22}^{-1} + \{\delta + (s_g - s_u)k - 1\}\gamma\Psi_{1|2}^{-1}\gamma' + l\Psi_{22}^{-1}.$$

We also obtain

$$E(\log|\Sigma_{gg}|\,|X^{(2)}=x^{(2)},\Phi_{gg},\delta)=-s_g k\log 2-\sum_{i=1}^{l}\Psi\left\{\frac{\delta+(s_g-s_u)k-i}{2}\right\}$$
$$-\sum_{i=1}^{s_g k-l}\Psi\left\{\frac{\delta+n+(s_g-s_u)k-l-h-i}{2}\right\}+\log|\Psi_{1|2}|+\log|\hat{\Psi}_{22}|,$$

where $\Psi(\cdot)$ is the digamma function, i.e. the first derivative of a gamma function.

To estimate the prior degrees of freedom, we simply need to solve the 'likelihood' equation

$$-s_g k\log 2-\log|\Sigma_{gg}|+\log|\Phi_{gg}|-\sum_{i=1}^{s_g k}\Psi\left\{\frac{\delta+(s_g-s_u)k-i}{2}\right\}=0.$$

By replacing $\log|\Sigma_{gg}|$ with $E(\log|\Sigma_{gg}||X^{(2)}=x^{(2)},\Phi_{gg},\delta)$ in this equation and using the relationship $\log|\Phi_{gg}|=\log|R'\Phi_{gg}R|=\log|\Psi_{1|2}|+\log|\Psi_{22}|$, the likelihood equation becomes

$$\sum_{i=1}^{s_g k-l}\Psi\left\{\frac{\delta+n+(s_g-s_u)k-l-h-i}{2}\right\}-\sum_{i=l+1}^{s_g k}\Psi\left\{\frac{\delta+(s_g-s_u)k-i}{2}\right\}$$
$$-\log|\hat{\Psi}_{22}|+\log|\Psi_{22}|=0$$

or equivalently

$$\sum_{i=l+1}^{s_g k}\left[\Psi\left\{\frac{\delta+n+(s_g-s_u)k-h-i}{2}\right\}-\Psi\left\{\frac{\delta+(s_g-s_u)k-i}{2}\right\}\right]$$
$$=\log|\hat{\Psi}_{22}|-\log|\Psi_{22}|.$$

Since the gamma function is convex, the digamma function is monotonic. Thus the left-hand side of the equation is always positive whenever $n-h>0$ and goes to 0 when $\delta$ increases to $\infty$. The right-hand side of the equation is positive. Therefore the solution of the above equation exists uniquely.

We can thus summarize our EM algorithm as follows.

(a) *E-step*: given the current values of $\Psi_{gg}$ and $\delta$,

$$E(\Sigma_{gg}^{-1}|X^{(2)}=x^{(2)})=R\begin{pmatrix}\{\delta+(s_g-s_u)k-1\}\Psi_{1|2}^{-1} & -\{\delta+(s_g-s_u)k-1\}\Psi_{1|2}^{-1}\gamma'\\ -\{\delta+(s_g-s_u)k-1\}\gamma\Psi_{1|2}^{-1} & d_1\end{pmatrix}R'$$

where

$$d_1=\{\delta+(s_g-s_u)k+n-l-h-1\}\hat{\Psi}_{22}^{-1}+\{\delta+(s_g-s_u)k-1\}\gamma\Psi_{1|2}^{-1}\gamma'+l\Psi_{22}^{-1}$$

and

$$E(\log|\Sigma_{gg}|\,|X^{(2)}=x^{(2)},\Phi_{gg},\delta)=-s_gk\log 2-\sum_{i=1}^{l}\Psi\left\{\frac{\delta+(s_g-s_u)k-i}{2}\right\}$$

$$-\sum_{i=1}^{s_gk-l}\Psi\left\{\frac{\delta+n+(s_g-s_u)k-l-h-i}{2}\right\}+\log|\Psi_{1|2}|+\log|\hat{\Psi}_{22}|.$$

(b) *M-step*: given the current values of $\Sigma_{gg}^{-1}$ and $\log|\Sigma_{gg}|$, find the maximum likelihood estimate of $\Phi_{gg}=\Lambda_g\otimes\Omega$ and $\delta$ by repeating the following steps (i) and (ii) until convergence and then go on to step (iii):

(i) given the current $\Lambda_g^{(p)}$ and $\Omega^{(p)}$, represent $\mathrm{tr}\{(\Lambda_g^{(p)}\otimes\Omega^{(p)})\Sigma_{gg}^{-1}\}$ as $\mathrm{tr}(\Omega^{(p)}G)$ and set $\Omega^{(p+1)}=s_g\{\delta+(s_g-s_u)k\}G^{-1}$;

(ii) given the current $\Lambda_g^{(p)}$ and $\Omega^{(p+1)}$, represent $\mathrm{tr}\{(\Lambda_g^{(p)}\otimes\Omega^{(p+1)})\Sigma_{gg}^{-1}\}$ as $\mathrm{tr}(\Lambda_g^{(p)}Q)$ and set $\Lambda_g^{(p+1)}=k\{\delta+(s_g-s_u)k-1\}Q^{-1}$;

(iii) given the current $\Psi_{gg}$, estimate $\delta$ by solving the equation

$$\sum_{i=l+1}^{s_gk}\left[\Psi\left\{\frac{\delta+n+(s_g-s_u)k-h-i}{2}\right\}-\Psi\left\{\frac{\delta+(s_g-s_u)k-i}{2}\right\}\right]=\log|\hat{\Psi}_{22}|-\log|\Psi_{22}|. \tag{8}$$

### 3.2. *Estimation of $B^0$ and $F^{-1}$*

Since $X^{(2)}=X^{(g)}R_2$,

$$X^{(2)}\sim ZB_gR_2+\mathcal{N}(I_n,\Sigma_{22}).$$

Define $\hat{B}_2=(Z'Z)^{-1}Z'X^{(2)}$ and $S_2=X^{(2)\prime}(I-Z(Z'Z)^{-1}Z')X^{(2)}$. Anderson (1984), p. 291, showed that, given $B_g$ and $\Sigma_{gg}$,

$$\hat{B}_2\sim B_gR_2+\mathcal{N}\{(Z'Z)^{-1},\Sigma_{22}\}, \tag{9}$$

$$S_2\sim W_{s_gk-l}(\Sigma_{22},n-h). \tag{10}$$

To estimate $B^0$ and $F^{-1}$, assume that $B^0$ has an exchangeable prior stochastic structure; i.e. that the columns of $B^0$ corresponding to the same item have the same mean. We denote this mean for item $i$ as $\mu_0^i$. We can readily find unbiased estimators of $\mu_0^i$ and $F^{-1}$. For this let

$$\hat{B}_2=\left(\hat{\beta}_1^{j_1},\ldots,\hat{\beta}_{s_gk-l}^{j_{s_gk-l}}\right),$$

where $j_v\in\{1,\ldots,k\}$, $v=1,\ldots,s_gk-l$. Obviously $(\hat{\beta}_v^{j_v})'=(Z'Z)^{-1}Z'X_v^{(2)}$ and $j_v$ marks the item type of $X_v^{(2)}$.

Let $\hat{\mu}_0^i$ be the mean of all $\hat{\beta}_v^{j_v}$ with $j_v=i$, $i=1,\ldots,k$, and

$$\hat{F}^{-1}=\frac{n-h-2}{s_gk-l}\sum_{v=1}^{s_gk-l}\frac{(\hat{\beta}_v^{j_v}-\hat{\mu}_0^{j_v})'(\hat{\beta}_v^{j_v}-\hat{\mu}_0^{j_v})}{a_{j_v}'S_2a_{j_v}}-(Z'Z)^{-1}$$

where $\hat{\beta}_v^{j_v}-\hat{\mu}_0^{j_v}=a_{j_v}'\hat{B}_2$. Then, given $X^{(2)}=x^{(2)}$ and the exchangeability of $B^0$, $\hat{\mu}_0^i$, $i=1,\ldots,k$, and $\hat{F}^{-1}$ are unbiased estimators of $\mu_0^i$, $i=1,\ldots,k$, and $F^{-1}$ respectively.

## 4. DISCUSSION AND CONCLUSIONS

The method developed in this paper can increase the effective number of gauged sites substantially, going from just seven to 31 for the application treated in Brown *et al.* (1994a). That gain can yield substantial increases in the accuracy of the interpolator as shown by a cross-validatory investigation reported in Sun *et al.* (1995). Also, we found the predictive distribution to be well calibrated so that, for example, nominal and actual coverage fractions for 95% posterior credibility intervals were approximately equal. Those results demonstrate anew the power of the hierarchical Bayesian approach and encourage further development of the basic methodology to expand its domain of applicability to include such things as binary space–time processes.

Our method unlike co-kriging uses past data and our empirical tests indicate that substantial gains can be made from using it. With our method comes an associated Bayesian design theory to be presented elsewhere. We also obtain a predictive distribution which can be coupled to impact models such as those used in environmental health analysis. Moreover, that distribution can be used for simulation studies.

Some technical improvements remain to be done. For example, it would be desirable to complete the hierarchical Bayesian model and to obtain a fully Bayesian method. Work in that direction is currently under way.

The use of the inverted Wishart distribution presents a limitation where only one parameter is allowed to reflect the uncertainty associated with the hypercovariance. As pointed out by a referee, this limitation may become critical in the missing by design situation where sites may measure different items and some sites may have very few observations compared with others. Hence more flexible prior distributions may be needed to escape the limitation. Brown *et al.* (1994b) have recently proposed a class of such prior distributions called *generalized inverted Wishart* (GIW) distributions. Work to incorporate the GIW prior distributions into the current methodology is under way.

## ACKNOWLEDGEMENTS

We are indebted to Dr Rick Burnett of Health Canada for pointing out the importance of the problem of interpolating fields with data missing by design. We thank the Joint Editor and reviewers for suggesting improvements to the quality of our presentation. Final revisions were completed while the last author was on leave at the Institute of Mathematics and Statistics, University of Kent at Canterbury, which generously provided facilities. At the same time he was partially supported by a Fellowship from the Engineering and Physical Sciences Research Council of the UK. The work reported in this paper was also partially supported by grants from the National Science and Engineering Council of Canada and by a contract with Health Canada.

## REFERENCES

Anderson, T. W. (1984) *An Introduction to Multivariate Statistical Analysis.* New York: Wiley.
Brown, P. J. (1993) *Measurement, Regression, and Calibration.* Oxford: Clarendon.

Brown, P. J., Le, N. D. and Zidek, J. V. (1994a) Multivariate spatial interpolation and exposure to air pollutants. *Can. J. Statist.*, **22**, 489–509.

———(1994b) Inference for a covariance matrix. In *Aspects of Uncertainty: a Tribute to D. V. Lindley* (eds A. F. M. Smith and P. R. Freeman). Chichester: Wiley.

Chen, C.-F. (1979) Bayesian inference for a normal dispersion matrix and its application to stochastic multiple regression analysis. *J. R. Statist. Soc.* B, **41**, 235–248.

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

Dawid, A. P. (1981) Some matrix–variate distribution theory. *Biometrika,* **68**, 265–274.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc.* B, **39**, 1–38.

Haas, T. C. (1990) Lognormal and moving window methods of estimating acid deposition. *J. Am. Statist. Ass.*, **85**, 950–963.

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

Press, S. J. (1982) *Applied Multivariate Analysis—using Bayesian and Frequentist Methods of Inference.* New York: Holt, Rinehart and Winston.

Sampson, P. and Guttorp, P. (1992) Nonparametric estimation of nonstationary spatial covariance structure. *J. Am. Statist. Ass.*, **87**, 108–119.

Sun, W. (1994) Bayesian multivariate interpolation with missing data and its applications. *PhD Thesis.* Department of Statistics, University of British Columbia, Vancouver.

Sun, W., Le, N. D., Zidek, J. V. and Burnett, R. (1995) Bayesian multivariate spatial interpolation: application and assessment. *Technical Report 146.* Department of Statistics, University of British Columbia, Vancouver.

Wu, C. F. J. (1983) On the convergence properties of the EM algorithm. *Ann. Statist.*, **11**, 95–103.