Bayesian Spatial Interpolation and Backcasting using Gaussian-Generalized Inverted Wishart Model

Nhu D Le, Li Sun, James V Zidek
University of British Columbia
June 14, 1999

Contents

Bayesian Spatial Interpolation and Backcasting using Gaussian- Generalized Inverted Wishart Model.

1 Introduction

Risk assessments of air pollution often require estimates of the concentration levels at locations where there are no monitoring sites using data available at monitored locations. Spatial interpolation to locations where people live using observed concentration levels at monitored sites is essential for such studies (c.f. Duddek et al. 1995, Zidek 1997, Zidek et al. 1998). For chronic diseases with long latency such as cancer where cumulative exposure is likely more relevant, the estimates for concentration levels are needed for long periods of time. Hence the observed concentration levels at different time periods are used for the estimation at other time periods. For example, the situation could arise in a case-control setting where residential histories of the participants are obtained and the estimates of concentration levels at the residential locations are needed for obtaining cumulative exposure estimates. The interpolation is sometimes needed at monitored locations for time periods before the installation of monitors. We call this interpolation situation *backcasting*.

In recent years, a Bayesian methodology for both temporal and spatial interpolation has been developed by Le and Zidek (1992) as an alternative to the well-known method of Kriging (c.f Cressie 1991). The method has been further enhanced to deal with the multivariate setting where possibly not all monitored sites measured the same set of pollutants (Brown, Le, Zidek 1994a and Le, Sun and Zidek 1997). The method produces the joint predictive distribution for several locations and different time points using all available data; thus allowing for simultaneous temporal and spatial interpolation. Another advantage of this method is that it does not assume the random field to be spatially isotropic. Furthermore, it allows for uncertainty associated with the mean and the spatial covariance of the field to be incorporated in the predictive distribution.

One important restriction in these methodological developments is that all monitored stations are required to be in operation for the same period. This restriction limits the flexibility of the methodology in that over time stations might be added to monitoring networks or data collected from networks with different operational periods could not be combined. In this work we consider the temporal and spatial interpolation problem where stations have been added over time to a monitoring network, creating a *staircase structure* of the observed data; that is, when the data are put together in an increasing order of the stations' operational periods, it shows a staircase structure. Similarly combining active networks with different starting times will yield the staircase structure as well. The staircase feature can be demonstrated with the air pollution data from monitoring networks maintained by the Ministry of Environment in the Province of British Columbia (BC). In figure 1, the monthly average ozone levels obtained from the 25 stations are displayed. The solid lines represent the observed monthly averages for stations with different starting times, showing the staircase structure.

**** Figure 1 here ****

In the paper, we develop a Bayesian method for temporal and spatial interpolation using all available data with this special feature. The method is based on the Gaussian and Generalized Inverted Wishart distribution. 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 (Brown, Le, Zidek 1994). The developed methodology yields the joint predictive distribution for several locations and time points using all staircase observed data. The method is an extension of the Bayesian developments discussed above and thus enjoys all the corresponding advantages. The method is applied to the above BC data. The mean of the predictive distribution for the backcasting component is displayed in Figure 1 (dash lines). Details will be given and discussed in later sections. We find that the method works reasonably well where the predicted means seem to be able to capture not only the temporal features demonstrated at specific stations but also the spatial feature of the pollution field.

The paper is organized as follows. The main theoretical results are described in Section 2 with relevant proofs given in the appendices. The parameter estimation is discussed in Section 3. The application of the method is illustrated in Section 4. Conclusion remarks then follow in Section 5.

2 Main Results

2.1 Notation

Throughout the paper, we adopt the notation used in Anderson (1984) when feasible. As well we let: etr $(\cdot) = \exp[\mathrm{tr}(\cdot)]$; $vec(X)$ be the $nm \times 1$ vector obtained by transposing and then stacking as columns the row vectors of X successively to form one column vector; $\otimes$ represent the Kronecker product between matrices. Additionally, let

$$\begin{aligned} n &= \text{number of time points (eg. number of months);} \\ u &= \text{number of locations with no monitors - called ungauged sites} \\ g &= \text{number of locations with monitors- called gauged sites.} \end{aligned}$$

The g gauged sites are organized into k blocks of stations where the g_j stations in the jth block have the same number of unobserved time points m_j. These blocks are numbered so that the observed measurements correspond to a staircase structure, that is,

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

To facilitate our presentation, we introduce the following notation for the response variables.

- Denote the response variables at the gauged and ungauged sites by

$$Y = \left[Y^{[u]}, \begin{pmatrix} Y^{[g_1^1]} \\ Y^{[g_1^2]} \end{pmatrix}, \ldots, \begin{pmatrix} Y^{[g_k^1]} \\ Y^{[g_k^2]} \end{pmatrix} \right],$$

  where:

  - $Y^{[u]}$, $n \times u$ matrix, denotes the responses at ungauged sites;
  - $Y^{[g_j^1]}$, $m_j \times g_j$ matrix, denotes the unobserved responses at the g_j gauged sites for the m_j time points;
  - $Y^{[g_j^2]}$, $(n - m_j) \times g_j$ matrix, denotes the responses at the g_j gauged sites with observations;

– $g = g_1 + \cdots + g_k$.

- observed measurements at the gauged sites where

$$D = \left\{ y^{[g_1^2]}, \cdots, y^{[g_k^2]} \right\}.$$

- unobserved responses by

$$Y_{unob} = \left\{ Y^{[u]}, Y^{[g_1^1]}, \cdots, Y^{[g_k^1]} \right\}.$$

- unobserved responses in blocks j to k by

$$Y^{[g_j^1, \ldots, g_k^1]} = \left\{ Y^{[g_j^1]}, \ldots, Y^{[g_k^1]} \right\}.$$

- responses from gauged blocks j to k, including both observed and unobserved stacks by

$$Y^{[g_j, \ldots, g_k]} = \left[\begin{pmatrix} Y^{[g_j^1]} \\ y^{[g_j^2]} \end{pmatrix}, \cdots, \begin{pmatrix} Y^{[g_k^1]} \\ y^{[g_k^2]} \end{pmatrix} \right].$$

- the responses from all gauged sites by

$$Y^{[g]} = Y^{[g_1, \ldots, g_k]}.$$

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

$$Z = \begin{pmatrix} Z_1' \\ \vdots \\ Z_n' \end{pmatrix}.$$

The coefficient matrix $\boldsymbol{\beta}$ of $l \times (u+g)$ corresponding to the l covariates and covariance matrix Σ of dimensioin $(u+g) \times (u+g)$ over gauged and ungauged sites are partitioned conformably:

$$\boldsymbol{\beta} = (\boldsymbol{\beta}^{[u]}, \boldsymbol{\beta}^{[g]}), \quad \Sigma = \begin{pmatrix} \Sigma^{[u]} & \Sigma^{[ug]} \\ \Sigma^{[gu]} & \Sigma^{[g]} \end{pmatrix}.$$

The coefficient matrix $\boldsymbol{\beta}^{[g]}$ and covariance matrix $\Sigma^{[g]}$ on the gauged sites are further partioned by blocks as

$$\boldsymbol{\beta}^{[g]} = (\boldsymbol{\beta}^{[g_1]}, \cdots, \boldsymbol{\beta}^{[g_k]}), \boldsymbol{\beta}^{[g_j, \ldots, g_k]} = (\boldsymbol{\beta}^{[g_j]}, \cdots, \boldsymbol{\beta}^{[g_k]})$$

and correspondingly

$$\Sigma^{[g]} = \begin{pmatrix} \Sigma^{[g_1]} & \ldots & \Sigma^{[g_1, g_k]} \\ \ldots & \ldots & \ldots \\ \Sigma^{[g_k, g_1]} & \ldots & \Sigma^{[g_k]} \end{pmatrix}, \quad \Sigma^{[g_j, \ldots, g_k]} = \begin{pmatrix} \Sigma^{[g_j]} & \ldots & \Sigma^{[g_j, g_k]} \\ \ldots & \ldots & \ldots \\ \Sigma^{[g_k, g_j]} & \ldots & \Sigma^{[g_k]} \end{pmatrix}.$$

The following 1-1 transformation (Barlett, 1933) of the matrix Σ is used:

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

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

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

where

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

for $j = 1, \ldots, k-1$. $\Sigma^{[g]}$ can then be obtained from $\{\Sigma_{kk}, (\Gamma_{k-1}, \tau_{k-1}), \ldots, (\Gamma_1, \tau_1)\}$ by means of this transformation. However the latter proves a much more convenient parameterization of our model.

2.2 The Model

The response matrix, Y, is assumed to follow the Gaussian-Generalized-Inverted-Wishart model specified by:

$$\begin{cases} Y \mid \boldsymbol{\beta}, \Sigma \sim N(Z\boldsymbol{\beta}, I_n \otimes \Sigma); \\ \boldsymbol{\beta} \mid \Sigma, \boldsymbol{\beta}_0, F \sim N(\boldsymbol{\beta}_0, F^{-1} \otimes \Sigma); \\ \Sigma \sim GIW(\Psi, \delta). \end{cases} \quad (1)$$

Here Z is the matrix of covariates and GIW represents the Generalized Inverted Wishart distribution developed by Brown, Le, Zidek (1994b). Specifically, $\Sigma \sim GIW(\Psi, \delta)$ is defined as:

$$\begin{cases} \Sigma^{[g]} \sim GIW(\Psi^{[g]}, \delta^{[g]}); \\ \Gamma^{[u]} = \Sigma^{[u|g]} = \Sigma^{[u]} - \Sigma^{[ug]}(\Sigma^{[g]})^{-1}\Sigma^{[gu]} \sim IW(\Psi^{[u]}, \delta^{[u]}); \\ \tau^{[u]} = (\Sigma^{[g]})^{-1}\Sigma^{[gu]} \sim N\left(\tau_0^{[u]}, H^{[u]} \otimes \Gamma^{[u]}\right) \end{cases} \quad (2)$$

where

$$\Gamma^{[u]} = \Sigma^{[u|g]} = \Sigma^{[u]} - \Sigma^{[ug]}(\Sigma^{[g]})^{-1}\Sigma^{[gu]}; \quad \tau^{[u]} = (\Sigma^{[g]})^{-1}\Sigma^{[gu]};$$

and IW denotes the Inverted Wishart distribution. The hypercovariance matrix Ψ is partitioned conformably with the partition of Σ.

The distribution of $\Sigma^{[g]}$ is further specified in terms of $\{\Sigma_{kk}, (\Gamma_{k-1}, \tau_{k-1}), \ldots, (\Gamma_1, \tau_1)\}$ as follows:

$$\begin{cases} \Sigma_{kk} \sim IW(\Psi_k, \delta_k); \\ \tau_j \mid \Gamma_j \sim N(\tau_{0j}, H_j \otimes \Gamma_j); \\ \Gamma_j \sim IW(\Psi_j, \delta_j) \end{cases} \quad (3)$$

for $j=1,\ldots,k-1$ and the δ's represent the degrees of freedom. The hyperparameters involved in the GIW model are written as

$$\mathcal{H}=\left\{\boldsymbol{\beta}_0, F, \Psi^{[u]}, \delta^{[u]}, \tau_0^{[u]}, H^{[u]}, \Psi_k, \delta_k, [\tau_{0j}, H_j, \Psi_j, \delta_j], j=1,\ldots,k-1\right\}, \tag{4}$$

The GIW distribution is a conjugate prior for Gaussian distributions. 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 each the k blocks can be expressed through the hyperparameter vector δ. More details on the characteristics of the GIW distribution are given in Brown, Le and Zidek (1994b).

2.3 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 gauge 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.

Dealing with the complex monotone data structures confronted in this paper forces us to pay a heavy notational price. Corresponding to the three parts of the theorem below, we introduce notation to facilitate its presentation.

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

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

$$= I_n + ZF^{-1}Z' + \tilde{\epsilon}^{[g_{j+1},\ldots,g_k]}H_j(\tilde{\epsilon}^{[g_{j+1},\ldots,g_k]})',$$

where

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

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

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

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

$$\Psi_{(1|2)}^{[j]} = \tfrac{\Psi_j}{\delta_j - g_1 + 1} + (\tilde{e}^{[g_j^2]})'(A_{22}^{[j]})^{-1}\tilde{e}^{[g_j^2]};$$

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

where

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

and

$$\mu^{[u|g]} = Z\boldsymbol{\beta}_0^{[u]} + \tilde{\epsilon}^{[g]}\tau_0^{[u]};$$

$$\Phi^{[u|g]} = I_n + ZF^{-1}Z' + \tilde{\epsilon}^{[g]}H^{[u]}(\tilde{\epsilon}^{[g]})';$$

where

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

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

$$(Y_{unob} \mid D, \mathcal{H}) \sim \left(Y^{[u]} \mid Y^{[g_1^1,\ldots,g_k^1]}, D, \mathcal{H}\right) \prod_{j=1}^{k-1} \left(Y^{[g_j^1]} \mid Y^{[g_{j+1}^1,\ldots,g_k^1]}, D, \mathcal{H}\right) \times \left(Y^{[g_k^1]} \mid D, \mathcal{H}\right), \quad (5)$$

where the three components of the conditional distributions are specified as following.

(i)

$$\left(Y^{[g_k^1]} \mid D, \mathcal{H}\right) \sim t_{m_k \times g_k} \left(\mu_{(1|2)}^{[k]}, \Phi_{(1|2)}^{[k]} \otimes \Psi_{(1|2)}^{[k]}, \delta_{(1|2)}^{[k]}\right). \quad (6)$$

(ii)

$$\left(Y^{[g_j^1]} \mid Y^{[g_{j+1}^1,\ldots,g_k^1]}, D, \mathcal{H}\right) \sim t_{m_j \times g_j} \left(\mu_{(1|2)}^{[j]}, \Phi_{(1|2)}^{[j]} \otimes \Psi_{(1|2)}^{[j]}, \delta_{(1|2)}^{[j]}\right). \quad (7)$$

(iii)

$$\left(Y^{[u]} \mid Y^{[g_1^1,\ldots,g_k^1]}, D, \mathcal{H}\right) \sim t_{n \times u} \left(\mu^{[u|g]}, (\delta^{[u]} - u + 1)^{-1}\Psi^{[u|g]} \otimes \Psi^{[u]}, \delta^{[u]} - u + 1\right). \quad (8)$$

Proof. See Appendix C.

We refer to the factors in (6) and (7) as *backcasting* since they give the joint predictive distribution of the response variables at the gauged sites during their ungauged time period. More precisely

$$(Y_{backcasting} \mid D) = \prod_{j=1}^{k-1} \left(Y^{[g_j^1]} \mid Y^{[g_j^2]}, Y^{[g_{j+1},\ldots,g_k]}\right) \times \left(Y^{[g_k^1]} \mid D, \mathcal{H}\right).$$

We refer to the factor in (8) 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

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

Let

$$\mu^{[g_j^1]} = E\left(Y^{[g_j^1]} \mid D, \mathcal{H}\right)$$

denote the predictive mean and

$$V^{[g_j^1]} = \text{var}\left(Y^{[g_j^1]} \mid D, \mathcal{H}\right),$$

the predictive variance of $Y^{[g_j^1]}$ given data and hyperparameters. For any matrix C, denote

$$C_{1,j} = \text{ the first } m_j \text{ rows of } C;$$

$$C_{2,j} = \text{ the last } n - m_j \text{ rows of } C.$$

Corollary 1: *The means of the predictive distributions in Theorem 1 are given below:*

$$\mu^{[g_k^1]} = \mu_{(1|2)}^{[k]},$$

$$\mu^{[g_j^1]} = E\left(\mu_{(1|2)}^{[j|j+1,\ldots,k]} \mid D, \mathcal{H}\right) = E\left(\mu_{(1)}^{[j]} \mid D, \mathcal{H}\right) + E\left(A_{12}^{[j]} \mid D, \mathcal{H}\right)(A_{22}^{[j]})^{-1}(Y^{[g_j^2]} - \mu_{(2)}^{[j]})$$

for $j = 1, \ldots, k-1$. Here $E\left(\mu_{(1)}^{[j]} \mid D, \mathcal{H}\right)$ and $E\left(A_{12}^{[j]} \mid D, \mathcal{H}\right)$ are recursively computed as

$$E\left(\mu_{(1)}^{[j]} \mid D, \mathcal{H}\right) = Z_{1,j}\boldsymbol{\beta}_0^{[g_j]} + \left(\mu_{1,j}^{[g_{j+1},\ldots,g_k]} - Z_{1,j}\boldsymbol{\beta}_0^{[g_{j+1},\ldots,g_k]}\right)\tau_{0j},$$

$$E\left(A_{12}^{[j]} \mid D, \mathcal{H}\right) = Z_{1,j}F^{-1}Z_{2,j} + \left(\mu_{1,j}^{[g_{j+1},\ldots,g_k]} - Z_{1,j}\boldsymbol{\beta}_0^{[g_{j+1},\ldots,g_k]}\right)H_j\left(Y_{2,j}^{[g_{j+1},\ldots,g_k]} - Z_{2,j}\boldsymbol{\beta}_0^{[g_{j+1},\ldots,g_k]}\right)'$$

where

$$\mu^{[g_{j+1},\ldots,g_k]} = \left[\begin{pmatrix} \mu^{[g_{j+1}^1]} \\ Y^{[g_{j+1}^2]} \end{pmatrix} \cdots \begin{pmatrix} \mu^{[g_k^1]} \\ Y^{[g_k^2]} \end{pmatrix}\right].$$

Proof. See Appendix C.

2.4 The Posterior Distributions

While our interest in this paper centers on predicting unobserved responses, we recognize that in some applications the model parameters themshelves will be of inferential interest. The model "transfer" coefficients in $\boldsymbol{\beta}$ 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 like wind speed or direction. Thus we develope and investigate the relevant posterior distributions in this section.

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

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

$$Z_{(j)} = K_j Z,$$

$$Y_{(j)}^{[g_{j+1},\ldots,g_k]} = K_j Y^{[g_{j+1},\ldots,g_k]},$$

$$\tilde{F}_j = Z_{(j)}Z'_{(j)} + F,$$

$$W_j = \tilde{F}_j^{-1}Z'_{(j)}Z_{(j)},$$

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

$$\hat{\boldsymbol{\beta}}^{[g_j^2]} = (Z_{(j)} Z'_{(j)})^{-1} Z'_{(j)} Y^{[g_j^2]},$$

$$\hat{\boldsymbol{\beta}}^{[g_{j+1},\ldots,g_k]} = (Z_{(j)} Z'_{(j)})^{-1} Z'_{(j)} Y_{(j)}^{[g_{j+1},\ldots,g_k]},$$

$$\tilde{\boldsymbol{\beta}}^{[g_j^2]} = W_j \hat{\boldsymbol{\beta}}^{[g_j^2]} + (I - W_j) \boldsymbol{\beta}_0^{[g_j]},$$

$$\tilde{\boldsymbol{\beta}}^{[g_{j+1},\ldots,g_k]} = W_j \hat{\boldsymbol{\beta}}^{[g_{j+1},\ldots,g_k]} + (I - W_j) \boldsymbol{\beta}_0^{[g_{j+1},\ldots,g_k]};$$

$$\tilde{\epsilon}^{[g_j^2]} = Y^{[g_j^2]} - Z_{(j)} \boldsymbol{\beta}_0^{[g_j]};$$

$$\tilde{\epsilon}_{(j)}^{[g_{j+1},\ldots,g_k]} = Y_{(j)}^{[g_{j+1},\ldots,g_k]} - Z_{(j)} \boldsymbol{\beta}_0^{[g_{j+1},\ldots,g_k]}.$$

Finally let

$$\tilde{\Psi}_k = \Psi_k + (\tilde{\epsilon}^{[g_k^2]})' [I_{n-m_k} + Z_{(k)} F^{-1} Z'_{(k)}]^{-1} \tilde{\epsilon}^{[g_k^2]},$$

$$\tilde{\Psi}_j = \Psi_j + (\tilde{\epsilon}^{[g_j^2]} - \tilde{\epsilon}_{(j)}^{[g_{j+1},\ldots,g_k]} \tau_{0j})'$$

$$\left[I_{n-m_j} + Z_{(j)} F^{-1} Z'_{(j)} + (\tilde{\epsilon}_{(j)}^{[g_{j+1},\ldots,g_k]})' H_j \tilde{\epsilon}_{(j)}^{[g_{j+1},\ldots,g_k]}\right] (\tilde{\epsilon}^{[g_j^2]} - \tilde{\epsilon}_{(j)}^{[g_{j+1},\ldots,g_k]} \tau_{0j});$$

$$\tilde{H}_j^{-1} = H_j^{-1} + (\tilde{\epsilon}_{(j)}^{[g_{j+1},\ldots,g_k]})' [I_{n-m_j} + Z_{(j)} F^{-1} Z'_{(j)}]^{-1} \tilde{\epsilon}_{(j)}^{[g_{j+1},\ldots,g_k]};$$

$$\tilde{\tau}_j = \tilde{H}_j \left[H_j^{-1} \tau_{0j} + (\tilde{\epsilon}_{(j)}^{[g_{j+1},\ldots,g_k]})' [I_{n-m_j} + Z_{(j)} F^{-1} Z'_{(j)}]^{-1} \tilde{\epsilon}^{[g_j^2]}\right];$$

$$\tilde{\delta}_j = \delta_j + n - m_j;$$

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

Theorem 2. *The joint posterior density is*

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

where

(i)

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

with

$$\begin{aligned} &\boldsymbol{\beta}^{[g_k]} \mid D, \Sigma_{kk}, \mathcal{H} \sim N\left(\tilde{\boldsymbol{\beta}}^{[g_k^2]}, \tilde{F}_k^{-1} \otimes \Sigma_{kk}\right); \\ &\boldsymbol{\beta}^{[g_j]} \mid D, \boldsymbol{\beta}^{[g_{j+1},\ldots,g_k]}, \tau_j, \Gamma_j, \mathcal{H} \sim N\left(\tilde{\boldsymbol{\beta}}^{[g_j^2]} + (\boldsymbol{\beta}^{[g_{j+1},\ldots,g_k]} - \tilde{\boldsymbol{\beta}}^{[g_{j+1},\ldots,g_k]}) \tau_j, \tilde{F}_j^{-1} \otimes \Gamma_j\right). \end{aligned} \tag{9}$$

(ii) $f(\Sigma \mid D, \mathcal{H})$ *is given by*

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

with

$$\begin{aligned} \Sigma_{kk} \mid D, \mathcal{H} &\sim IW(\tilde{\Psi}_k, \tilde{\delta}_k), \\ \tau_j \mid D, \Gamma_j, \mathcal{H} &\sim N\left(\tilde{\tau}_j, \tilde{H}_j \otimes \Gamma_j\right); \\ \Gamma_j \mid D, \mathcal{H} &\sim IW(\tilde{\Psi}_j, \tilde{\delta}_j). \end{aligned} \tag{10}$$

Proof. See Appendix C.

Corollary 2: *The posterior means of $\boldsymbol{\beta}$ given data D and hyperparameter set $\mathcal{H}$ are given below:*

$$E(\boldsymbol{\beta}^{[g_k]} \mid D, \mathcal{H}) = \tilde{\boldsymbol{\beta}}^{[g_k^2]},$$

$$E(\boldsymbol{\beta}^{[g_j]} \mid D, \mathcal{H}) = \tilde{\boldsymbol{\beta}}^{[g_j^2]} + \left[E(\boldsymbol{\beta}^{[g_{j+1},\ldots,g_k]} \mid D, \mathcal{H}) - \tilde{\boldsymbol{\beta}}^{[g_{j+1},\ldots,g_k]}\right] \tilde{\tau}_j,$$

where

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

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

Proof. See Appendix C.

We note that $E(\boldsymbol{\beta}^{[g_j]} \mid D, \mathcal{H}) = \tilde{\boldsymbol{\beta}}^{[g_j^2]}$ if data are complete, i.e., 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 D.

3 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 and Guttorp (1992).

3.1 Estimating Hyperparameters Using 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 resulted estimators are the Type II maximum likelihood estimates. This approach has been used in previous studies (Brown, Le and Zidek 1994a; Sun, Le, Zidek and 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 Barlett transformation. That is, with the recursive notation

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

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

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

In this work, we assume

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

It is easy to show that

$$\Psi_{(jj)} = \begin{pmatrix} \Psi_j + \tau'_{0j} \Psi_{(j+1,j+1)} \tau_{0j} & \tau'_{0j} \Psi_{(j+1,j+1)} \\ \Psi_{(j+1,j+1)} \tau_{0j} & \Psi_{(j+1,j+1)} \end{pmatrix}, \tag{11}$$

and that

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

Furthermore, we assume that the degrees of freedoms, $\delta_1, \ldots, \delta_k$, follow a gamma prior distribution (Le, Sun and Zidek, 1998) where

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

with α and r specified.

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

$$\begin{aligned} Q(\mathcal{H} \mid \mathcal{H}^{(p)}) &= E\left(\log[f(Y, \boldsymbol{\beta}, \Sigma \mid \mathcal{H})\pi(\delta)] \mid D, \mathcal{H}^{(p)}\right) \\ &= E\left[\log f(Y \mid \boldsymbol{\beta}, \Sigma) \mid D, \mathcal{H}^{(p)}\right] + E\left[\log f(\boldsymbol{\beta}, \Sigma \mid \mathcal{H}) \mid D, \mathcal{H}^{(p)}\right] + \log \pi(\delta) \end{aligned} \tag{12}$$

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 $\boldsymbol{\beta}$ and Σ with respect to the the posterior distribution $\boldsymbol{\beta}, \Sigma \mid D, \mathcal{H}^{(p)}$.

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

$$Q^*(\mathcal{H} \mid \mathcal{H}^{(p)}) = E\left[\log f(\boldsymbol{\beta}, \Sigma \mid \mathcal{H}) \mid D, \mathcal{H}^{(p)}\right] + \log \pi(\delta) \tag{13}$$

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

With the parameterization introduced above

$$\begin{aligned} f(\boldsymbol{\beta}, \Sigma \mid \mathcal{H}) &\propto \left(\textstyle\prod_{j=1}^k c(g_j, \delta_j)\right) |\Sigma_{kk}|^{-\frac{l+\delta_k+g_k+1}{2}} \left(\textstyle\prod_{j=1}^{k-1} |\Gamma_j|^{-\frac{l+\delta_j+g_j+\cdots+g_k+1}{2}}\right) \\ &\times \left(\textstyle\prod_{j=1}^k |\Psi_j|^{\frac{\delta_j+g_1+\cdots+g_{j-1}}{2}}\right) |F|^{\frac{g}{2}} \mathrm{etr}\left\{-\tfrac{1}{2}\Sigma^{-1}[\Psi + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' F (\boldsymbol{\beta} - \boldsymbol{\beta}_0)]\right\}. \end{aligned} \tag{14}$$

For simplicity assume $g_1 + g_0 = 0$ so that the power of Ψ_1 is $\delta_1/2$. Hence,

$$
\begin{aligned}
Q^*(\mathcal{H} \mid \mathcal{H}^{(p)}) &= E\left[\log f(\boldsymbol{\beta}, \Sigma \mid \mathcal{H}) \mid D, \mathcal{H}^{(p)}\right] + \log \pi(\delta) \\
&= const + \textstyle\sum_{j=1}^{k} \log c(g_j, \delta_j) + \log \pi(\delta) \\
&- \tfrac{1}{2}(l + \delta_k + g_k + 1) E\left[\log |\Sigma_{kk}| \mid D, \mathcal{H}^{(p)}\right] \\
&- \tfrac{1}{2} \textstyle\sum_{j=1}^{k-1} (l + \delta_j + g_j + \cdots + g_k + 1)) E\left[\log |\Gamma_j| \mid D, \mathcal{H}^{(p)}\right] \\
&+ \tfrac{1}{2} \textstyle\sum_{j=1}^{k} (\delta_j + g_1 + \cdots + g_{j-1}) \log |\Psi_j| + \tfrac{g}{2} \log |F| \\
&- \tfrac{1}{2} \mathrm{tr} E\left[\Sigma^{-1}[\Psi + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' F (\boldsymbol{\beta} - \boldsymbol{\beta}_0)] \mid D, \mathcal{H}^{(p)}\right].
\end{aligned}
\tag{15}
$$

Suppose the current estimate is

$$
\mathcal{H}^{(p)} = \left(\boldsymbol{\beta}_0^{(p)}, F^{(p)}, \Psi_k^{(p)}, \delta_k^{(p)}, [\tau_{0j}^{(p)}, \Psi_j^{(p)}, \delta_j^{(p)}], j = 1, \ldots, k-1\right). \tag{16}
$$

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

(i) E-step: Compute the posterior expectations involved in (15), given data and $\mathcal{H}^{(p)}$. We present the detailed calculations in the Appendix.

(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

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

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

(b) To update the estimates of Ψ, maximize

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

(c) To update the estimates of $\boldsymbol{\beta}_0$ and F, maximize

$$
g \log |F| - \mathrm{tr}\left\{F E\left[(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \Sigma^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \mid D, \mathcal{H}^{(p)}\right]\right\}.
$$

[Note that the estimation of (δ, Ψ) can be separated from that of $(\boldsymbol{\beta}, F)$].

To complete (a)-(c), consider the following recursive partition of $E\left[\Sigma^{-1} \mid D, \mathcal{H}^{(p)}\right]$

$$
\begin{aligned}
&E\left[\Sigma^{-1} \mid D, \mathcal{H}^{(p)}\right] \\
&= \begin{pmatrix} B_{11}^{(1)} & B_{12}^{(1)} \\ B_{21}^{(1)} & B_{22}^{(1)} \end{pmatrix} : \begin{pmatrix} g_1 \times g_1 & g_1 \times (g_2 + \cdots + g_k) \\ (g_2 + \cdots + g_k) \times g_1 & (g_2 + \cdots + g_k) \times (g_2 + \cdots + g_k) \end{pmatrix}
\end{aligned}
$$

and

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

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

for $j = 2, \ldots, k-1$. (Note that the partitions depend also on p, which is dropped from the index for simplicity).

It follows that the estimates of δ and Ψ can be iteratively updated by solving the equations

$$\frac{1}{2}\sum_{i=1}^{g_j}\left[\eta\left(\frac{\tilde{\delta}_j^{(p)} - i + 1}{2}\right) - \eta\left(\frac{\delta_j^{(p+1)} - i + 1}{2}\right)\right] + \frac{1}{2}\log\left|\Psi_j^{(p+1)}\right| - \frac{1}{2}\log\left|\tilde{\Psi}_j^{(p)}\right| + \frac{\alpha - 1}{\delta_j} - r = 0 \quad (17)$$

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

$$\Psi_j^{(p+1)} = (\delta_j^{(p+1)} + g_1 + \cdots + g_{j-1})(B_{11}^{(j)})^{-1}, \ \tau_{0j}^{(p+1)} = -B_{21}^{(j)}(B_{11}^{(j)})^{-1}, \quad (18)$$

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

$$\Psi_k^{(p+1)} = B_{22}^{(k-1)} - B_{21}^{(k-1)}(B_{11}^{(k-1)})^{-1}B_{12}^{(k-1)}, \quad (19)$$

(recall that $g_1 + g_0 = 0$), where $\tilde{\delta}_j^{(p)}$ and $\tilde{\Psi}_j^{(p)}$ are as given in the Appendix with parameters replaced by the current estimates at iteration p.

The updated version of estimates of $\boldsymbol{\beta}$ and F are, respectively

$$\boldsymbol{\beta}_0^{(p+1)} = \left(E[\Sigma^{-1} \mid D, \mathcal{H}^{(p)}]\right)^{-1} E[\Sigma^{-1}\boldsymbol{\beta} \mid D, \mathcal{H}^{(p)}] \quad (20)$$

and

$$F^{(p+1)} = g^{-1}\left\{E[\boldsymbol{\beta}\Sigma^{-1}\boldsymbol{\beta}' \mid D, \mathcal{H}^{(p)}] - E[\boldsymbol{\beta}\Sigma^{-1} \mid D, \mathcal{H}^{(p)}]\left(E[\Sigma^{-1} \mid D, \mathcal{H}^{(p)}]\right)^{-1} E\left[\Sigma^{-1}\boldsymbol{\beta}' \mid D, \mathcal{H}^{(p)}\right]\right\} \quad (21)$$

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 Barlett 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.

3.2 Extending the Covariance Matrix Estimate via the Sampson-Guttorp Method

In this subsection, 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. Hence 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 coverting the resulted spatial hypecovariance 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

$$M = \begin{pmatrix} M_{uu} & M_{ug} \\ M_{gu} & M_{gg} \end{pmatrix}. \tag{22}$$

Thus,

$$\tau_0^{[u]} = M_{gg}^{-1} M_{gu} \tag{23}$$

and

$$\Psi^{[u]} = \frac{(\delta_1 - u - 1)(\delta^{[u]} - g - 1)}{\delta^{[u]} - 1}(M_{uu} - M_{ug} M_{gg}^{-1} M_{gu}). \tag{24}$$

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}. More details on that method are given in the Application section.

The parameters of interest are estimated through equations (23) and (24) using the estimated M_{uu}, M_{gu} and M_{ug}.

4 Applications

We now demonstrate our model using BC ozone data, obtained from BC Ministry of Environment.

4.1 Data Set

The data set consists of monthly ozone levels over 25 monitor sites in the Province of British Columbia. The monthly levels were obtained 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 204 months from as early as January, 1978 up to December, 1994. Many sites started operation later than 1978, resulting in the stair-case data set pattern.

The locations of these sites are shown in Figure 2, their information about them is summarized in Table 1.

(Figure 2. is about here)

The latitude and longitude in Table 1 are in `<ddmmss>` format. For example, Abbotsford Library has latitude $49^o 03' 00''$ and longitude $122^o 19' 00''$.

The boxplot of the monthly average ozone levels of these 25 sites is shown in Figure 3. The means and standard deviations for these stations are given in Table 2. The unit of measurement is *ppb* (parts per billion).

(Figure 3. is about here)

Table 1: BC Ozone Monitor Sites

Site No.	Site ID	Latitude	Longitude	Location	Start
1	E217029	490300	1221900	ABBOTSFORD LIBRARY	92/2
2	E214615	492236	1212625	HOPE FIRE HALL	91/5
3	E209177	491930	1230500	MAHON PARK - NORTH VANCOUVER	90/4
4	E207418	491300	1225900	BURNABY SOUTH	87/10
5	E207421	491636	1224400	PORT COQUITLAM	87/6
6	E207723	490952	1225648	NORTH DELTA	87/6
7	E207417	490834	1230621	RICHMOND SOUTH	86/7
8	E206612	490909	1215703	CHILLIWACK	86/4
9	E206169	491300	1224300	PITT MEADOWS	85/5
10	E231001	492524	1225812	SEYMOUR FALL	85/5
11	E206270	491645	1225442	BURNABY MOUNTAIN	84/6
12	E206271	490804	1224136	SURREY EAST	84/1
13	X500886	495140	1192830	KELOWNA	83/8
14	X310162	491640	1225050	ROCKY POINT PARK - PORTMOODY	80/8
15	X310167	491840	1225130	ANMORE	80/8
16	X310169	491704	1224924	EAGLE RIDGE - PORTMOODY	80/8
17	X310174	491655	1230714	ROBERSON SQUARE - VANCOUVER	80/8
18	X310175	491540	1230920	KITSILANO - VANCOUVER	80/8
19	X310176	491230	1230650	MARPOLE - VANCOUVER	80/8
20	X310177	491640	1225810	KENSINGTON PARK - BURNABY	80/8
21	X310178	491700	1230000	CONFEDERATION PARK - BURNABY	80/8
22	X310179	491700	1230110	SECOND NARROWS	80/8
23	X310180	491900	1230800	WEST VANCOUVER	80/8
24	X110030	482530	1232130	VICTORIA	78/1
25	X310081	490150	1222234	ABBOTSFORD AIRPORT	78/1

4.2 Trend Modelling

Following a similar analysis of BLZ (1994), the covariate variables in our general theory for the response variable (the logorithm transformed monthly average ozone level) are $[1, \cos(2\pi t/12), \sin(2\pi t/12)]$.

Figure 4. shows the fits of the trend model to the original data for a typical site (Site 16). Similar results for other sites are omitted for brevity.

(Figure 4. is about here)

The residuals after detrending the log-transformed ozone concentration data showed no clear pattern, suggesting the Gaussian assumption required in our theory is to be approximately valid.

Our theory assumes no values are missing after a site started operation. Therefore, to proceed to GIW modelling, we need to impute randomly missing values for each site. We do this using the EM algorithm.

Table 2: Means And Standard Deviations of Ozone Levels

Site No.	1	2	3	4	5	6	7	8	9
Mean	25.88	18.29	24.82	23.00	26.27	24.92	25.75	28.55	28.43
SD	9.61	12.35	9.33	10.17	9.67	9.00	10.21	11.76	11.23

Site No.	10	11	12	13	14	15	16	17	18
Mean	35.56	38.16	34.20	27.53	22.54	32.12	22.39	9.041	19.05
SD	12.46	9.89	12.36	16.74	11.89	11.21	10.55	5.04	10.20

Site No.	19	20	21	22	23	24	25
Mean	19.80	22.91	27.26	20.24	21.07	25.93	32.80
SD	9.8	8.57	9.86	8.67	9.67	13.18	12.57

Let Y_{it} denote the logarithm of ozone level of site i at month t. We assume a structural model for Y_{it}, namely

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

where ϵ_{it}, $i = 1, \ldots, 25$ and $t = 1, \ldots, 204$, are independent over time with mean zero and a normal distribution. Assume the covariance of $(\epsilon_{1t}, \ldots, \epsilon_{25,t})$ is a 25×25 positive definite matrix V.

We first use the observed values at site i to estimate β_{i0} β_{i1} and β_{i2} by linear regression.

Given a positive definite matrix V_0 as an initial estimate of V, all missing values at time t are imputed as the conditional mean values according to the spatial correlation specified by V and standard normal theory.

The estimte of matrix V is then updated as the sample variance matrix after missing values are fitted.

The above two-step process is then iterated until convergence.

4.3 Spatial Interpolation and Backcasting

The Generalized Inverted Wishart modelling method we have developed 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. Or we could select each site as a block in the stair-case structure. The latter case includes flat stairs.

We apply our model to the ozone data with one site per block. Thus, we have $k = 25$ blocks with each block containing $g_1 = 1$ site. The total number of the gauged sites is $g = g_1 + \ldots + g_k = 25$.

As in the previous sub-section, the logarithm of the ozone level is denoted by Y_{it} for site i at month t and the covariate vector is $Z_t = [1, \cos(2\pi t/12), \sin(2\pi t/12)]$.

Only two out of twenty-five sites provided complete data from January, 1978 to December, 1994. They are Victoria (Site 24) and Abbotsford Airport (Site 25). The latest site is Abbotsford Library (Site 1) which started in February, 1992, so it lacks data corresponding to more than 14 years.

The model in Section 2.2 specifies the distributions of data Y, random effects vector $\boldsymbol{\beta}$ and the covariance matrix Σ. We assume the H matrix satisfies (11), so that the number of parameters in this example is $3(\boldsymbol{\beta}_0) + 6(F) + 25(\Psi_j) + 300(\tau_{0j}) + 25(\delta_j) = 359$.

The parameter estimates are obtained via the EM algorithm. The estimate of $\boldsymbol{\beta}_0$ is given by (3.06, -0.47, 0.11) and the estimated F matrix is

$$\begin{pmatrix} 1.07 & -2.30 & 2.82 \\ -2.30 & 6.80 & -7.48 \\ 2.82 & -7.48 & 13.26 \end{pmatrix}.$$

Using the results of Section 5.2, we impute the systematically missing values for the gauged sites. This imputation technique is called "backcasting". Figure 1 as presented in the Introduction section shows the backcasted ozone levels (dotted lines) along with the observed levels (solid lines).

The next phrase of our analysis is conducted using the SG methodology for interpolating the covariance structure over the region under study. The method develops a smooth mapping between G-space (geographic space) and D-space (dispersion space). The results of fitting the variogram and D-space coodinates by means of the spline smoothing technique are presented in Figure 5 and 6.

(Figure 5. is about here)

(Figure 6. is about here)

To enable examination of the transformation of the G-space, we present in Figure 7 the biorthogonal grid. Solid lines show directions in which contraction of the G-space is required, indicating the direction of high correlation. In contrast, the dotted lines indicate direction of stretching while the broken lines indicate severe stretching, corresponding to the directions of small correlation and very small correlation, respectively. We believe the west-east directions of strong correlation (solid line) are caused by the wind patterns in the Lower Mainland of BC. The very small correlation in the lower right corner is due to a relative lack of information.

(Figure 7. is about here)

In implementating the SG method, the transformation between the G-space and D-space is specified so that correlation between any two G-locations can be determined. We first tranform the coordinates of two locations in G-space to D-space and then compute the isotropic correlation between the tranformed locations in D-space by the fitted variogram function.

In our case, we consider 400 ungauged sites that are determined by a 20 by 20 equally spaced grid that covers the rectangular region determined by the 25 sites. Following the discussion in the previous section, the interpolated covariance matrix $M^{[a]}$ is first obtained via the SG method. We are then able to get the interpolated ozone levels for 400 ungauged sites from 1978 to 1994. This yields a 204×400 matrix of interpolants.

The following two figures show the interpolated distribution field of ozone in 1994.

(Figure 8. is about here)

(Figure 9. is about here)

The standard deviation for the interpolation of August 1994 is shown in Figure 10.

(Figure 10. is about here)

The contour plot shows that an ungauged site close to an existing (gauged) site has a small SD. Therefore, those ungauged sites with average interpolated ozone levels in the center of the region and on the north-west and south-east corners have bigger SD values. (In the original scale, the SD's are computed via the delta method).

5 Conclusion 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 Baysian methodology for spatial interpolation developed earlier by Le and Zidek (1992) which was further extended to multivariate setting by Brown, Le and Zidek (1994), and Le, Sun and Zidek (1997). This approach would allow for combining data from active networks with different starting times of operation in the interpolation and thus providing a 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 cross-validation setting.

The development of the interpolation approach comes with a related Bayesian design theory. 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 and 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.

Although monthly averages of ozone concentration levels do not exhibit temporal correlations, hourly or daily levels of various pollutants including ozone and PM10 do indicate strong temporal correlations. Spatial interpolation for these hourly or daily fields is quite crucial for examining the acute health impact of air pollution. Extension of the current method to take temporal correlations into account is under consideration.

Aknowledgement

This research is partially supported by grants from National Science and Engineering Research Council and BC Health Research Foundation. We are indebted to Mr. Ernie Tradewell of BC Ministry of Environment, Land and Parks for providing the data.

Appendices

A Multivariate and Matrix Distributions.

In this section of the Appendix we present for completeness some of the distributions commonly encountered in multivariate Gaussian theory. In our presentation we will let $\sim$ mean "distributed as". More generally, $U|V \sim$ will mean "conditional on V, U is distributed as". Because of its widespread usage in multivariate analysis, we adopt the notation of Anderson (1978) where feasible.

A.1 Matrix Normal Distribution.

We say $X_{n\times m} \sim N(X^{(0)}, A \otimes B)$ if X has the joint density

$$f(X) = (2\pi)^{-nm/2}|A|^{-m/2}|B|^{-n/2}\mathrm{etr}\left\{-\frac{1}{2}[A^{-1}(X - X^{(0)})][(X - X^{(0)})B^{-1}]'\right\}. \tag{25}$$

We note that $X \sim N(X^{(0)}, A \otimes B)$ if and only if $X' \sim N(X^{(0)'}, B \otimes A)$. Let

$$X = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix} = (x^{(1)}, \ldots, x^{(m)}), \tag{26}$$

and

$$A = (a_{ij})_{n\times n}, \quad B = (b_{ij})_{m\times m}.$$

Proposition 1. *Some properties of the matrix normal distribution are given below.*

- $E(X) = X^{(0)}$, $\mathrm{var}[vec(X)] = A \otimes B$, *and* $\mathrm{var}[vec(X')] = B \otimes A$.
- $\mathrm{cov}(x_i, x_j) = a_{ij}B, \quad \mathrm{cov}(x^{(i)}, x^{(j)}) = b_{ij}A.$
- *For any matrix* C*:* $c \times n$ *and matrix* D*:* $m \times d$,

$$CXD \sim N(CX^{(0)}D, CAC' \otimes D'BD).$$

- *For any* $m \times m$ *matrix* C *and* $n \times n$ *matrix* D,

$$EXCX' = X^{(0)}CX^{(0)'} + A\mathrm{tr}(CB), \quad EX'DX = X^{(0)'}DX^{(0)} + \mathrm{tr}(AD)B. \tag{27}$$

  In particular,

$$EXB^{-1}X' = X^{(0)}B^{-1}X^{(0)'} + mA, \quad EX'A^{-1}X = X^{(0)'}A^{-1}X^{(0)} + nB.$$

A.2 Multivariate and Matric Student t-distribution

We say a random vector Y has a p-dimensional student t-distribution with δ degrees of freedom, mean vector μ_0 and precision matrix H, or more simply $Y \sim t_p(\mu_0, H, \delta)$ if it has a density function of the form

$$f(Y) = \frac{\Gamma\left(\frac{\delta+p}{2}\right)|H|^{-\frac{1}{2}}}{\Gamma\left(\frac{\delta}{2}\right)(\delta\pi)^{\frac{p}{2}}}\left[1 + \frac{1}{\delta}(Y - \mu_0)'H^{-1}(Y - \mu_0)\right]^{-\frac{\delta+p}{2}}.$$

If $Y \sim t_p(\mu_0, H, \delta)$, then $E(Y) = \mu_0$ and $\text{var}(Y) = \delta H/(\delta - 2)$.

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

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

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

$$f(X) \propto |A|^{-m/2}|B|^{-n/2}|I_n + \delta^{-1}[A^{-1}(X - X^{(0)})][(X - X^{(0)})B^{-1}]'|^{-\frac{\delta+n+m-1}{2}}, \tag{28}$$

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

$$\Gamma_p(t) = \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma[t - (i-1)/2] \tag{29}$$

denotes the multivariate Gamma function.

Proposition 2. *Some properties of the matric t-distribution are given below.*

- $X \sim t_{n\times m}(X^{(0)},\ A \otimes B,\ \delta)$, *if and only if* $X' \sim t_{m\times n}(X^{(0)'},\ B \otimes A,\ \delta)$.
- *If* $n = 1$ *and* $A = 1$, X *has an* m*-variate* t*-distribution,* i.e.,

$$X \sim t_m(X^{(0)},\ B,\ \delta).$$

- *If* $m = 1$ *and* $B = 1$, X *has an* n*-variate* t*-distribution,* i.e.,

$$X \sim t_n(X^{(0)},\ A,\ \delta).$$

- *When* $\delta > 2$, $E(X) = X^{(0)}$ *and* $\text{var}[vec(X)] = \delta(\delta - 2)^{-1}A \otimes B$. *In particular,*

$$\text{cov}(x_i, x_j) = \delta(\delta - 2)^{-1}a_{ij}B, \quad \text{cov}(x^{(i)}, x^{(j)}) = \delta(\delta - 2)^{-1}b_{ij}A.$$

- *If* $X \sim t_{n\times m}(X^{(0)},\ A \otimes B,\ \delta)$, *and* $C : c \times n$ *is of rank* c *and* $D : m \times d$ *of rank* d, *then*

$$Y = CXD \sim t_{c\times d}(CX^{(0)}D,\ CAC' \otimes D'BD,\ \delta).$$

  In particular,

$$x_i \sim t_m(x_i^{(0)},\ a_{ii}B,\ \delta),$$

  and

$$x^{(j)} \sim t_n(x^{(0,j)},\ b_{jj}A,\ \delta),$$

  assuming $X^{(0)}$ *is formed like* X *as in (26).*

We give the following result on marginal and conditional matric t-distributions.

Proposition 3. *(a) Assume* $Y = t_{n\times m}(\mu, A\otimes B, \delta)$, *where*

$$Y = (Y^{[1]}, Y^{[2]}) : (n\times m_1, n\times m_2),$$

$$\mu = (\mu^{[1]}, \mu^{[2]}) : (n\times m_1, n\times m_2),$$

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} : \begin{pmatrix} m_1\times m_1 & m_1\times m_2 \\ m_2\times m_1 & m_2\times m_2 \end{pmatrix}.$$

We then have:

- *the marginal distribution*
$$Y^{[2]} \sim t_{n\times m_2}(\mu^{[2]}, A\otimes B_{22}, \delta);$$
- *the conditional distribution*
$$Y^{[1]} \mid Y^{[2]} \sim t_{n\times m_1}(\mu^{[1|2]}, A^{[1|2]}\otimes B^{[1|2]}, \delta + m_2)$$

  where
$$\mu^{[1|2]} = \mu^{[1]} + \tilde{\epsilon}^{[2]} B_{22}^{-1} B_{21};$$
$$A^{[1|2]} = A + \tilde{\epsilon}^{[2]} B_{22}^{-1} (\tilde{\epsilon}^{[2]})';$$
$$B^{[1|2]} = \tfrac{\delta}{\delta+m_2}(B_{11} - B_{12}B_{22}^{-1}B_{21});$$
$$\tilde{\epsilon}^{[2]} = Y^{[2]} - \mu^{[2]}.$$

(b) If $Y = t_{n\times m}(\mu, A\otimes B, \delta)$, *where*

$$Y = \begin{pmatrix} Y_{(1)} \\ Y_{(2)} \end{pmatrix} : \begin{pmatrix} n_1\times m \\ n_2\times m \end{pmatrix}$$

$$\mu = \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix} : \begin{pmatrix} n_1\times m \\ n_2\times m \end{pmatrix},$$

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} : \begin{pmatrix} n_1\times n_1 & n_1\times n_2 \\ n_2\times n_1 & n_2\times n_2 \end{pmatrix},$$

then we have

- *the marginal distribution*
$$Y_{(2)} \sim t_{n_2\times m}(\mu_{(2)}, A_{22}\otimes B, \delta),$$
- *the conditional distribution*
$$Y_{(1)} \mid Y_{(2)} \sim t_{n_1\times m}(\mu_{(1|2)}, A_{(1|2)}\otimes B_{(1|2)}, \delta + n_2),$$

where

$$\mu_{(1|2)} = \mu_{(1)} + A_{12}A_{22}^{-1}\tilde{\epsilon}_{(2)};$$

$$A_{(1|2)} = \tfrac{\delta}{\delta+n_2}(A_{11} - A_{12}A_{22}^{-1}A_{21});$$

$$B_{(1|2)} = B + (\tilde{\epsilon}_{(2)})'A_{22}^{-1}\tilde{\epsilon}_{(2)};$$

$$\tilde{\epsilon}_{(2)} = Y_{(2)} - \mu_{(2)}$$

A.3 Wishart and Inverted Wishart Distribution

We say that a $p \times p$ matrix Y has a Wishart distribution $W_p(\Sigma, \delta)$, if its joint density has the form:

$$f(Y) = c(p,\delta)|\Sigma|^{-\frac{1}{2}\delta}|Y|^{\frac{1}{2}(\delta-p-1)}\mathrm{etr}\left\{-\frac{1}{2}\Sigma^{-1}Y\right\} \tag{30}$$

where

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

We say that a $p \times p$ dimensional random matrix Y has an inverted Wishart distribution $Y \sim IW_p(\Psi, \delta)$ or $IW(\Psi, \delta)$ with covariance parameter Ψ and degrees of freedom δ, if Y has a joint density of the form

$$f(Y) = c(p,\delta)|\Psi|^{\frac{1}{2}\delta}|Y|^{-\frac{1}{2}(\delta+p+1)}\mathrm{etr}\left\{-\frac{1}{2}\Psi Y^{-1}\right\}. \tag{31}$$

When $p = 1$ the inverted Wishart distribution becomes inverted chi-squared distribution *i.e.* $IW_1(\psi, \delta) = \psi\chi_\delta^{-2}$ where $Y \sim \psi\chi_\delta^{-2}$ means that Y has the joint density

$$f(y) = \left[2^{\frac{\delta}{2}}\Gamma\left(\frac{\delta}{2}\right)\right]^{-1}\psi^{\frac{\delta}{2}}y^{-\frac{\delta+2}{2}}\exp\left\{-\frac{\psi}{2y}\right\}.$$

Proposition 4. *Some properties of the Wishart and Inverted Wishart distributions needed in the sequel are:*

- $Y \sim IW_p(\Psi, \delta)$ *iff* $Z = Y^{-1} \sim W_p(\Psi^{-1}, \delta)$.
- *If* $Z \sim W_p(\Sigma, \delta)$ *then* $E(Z) = \delta\Sigma$ *and* $E(Z^{-1}) = \Sigma^{-1}/(\delta - p - 1)$ *provided* $\delta - p - 1 > 0$.
- *If* $Y \sim IW_p(\Psi, \delta)$, *then* $E(Y) = \Psi/(\delta - p - 1)$ *and* $E(Y^{-1}) = \delta\Psi^{-1}$.
- *If* $Y \sim IW_p(\Psi, \delta)$, *then*

$$E\log|Y| = -p\log 2 - \sum_{i=1}^{p}\eta\left[\frac{1}{2}(\delta - i + 1)\right] + \log|\Psi| \tag{32}$$

where η *is the digamma function, that is,* $\eta(x) = d[\log\Gamma(x)]/dx$, *(see Chen 1979).*

B Simple Gaussian Inverted Wishart Model and Gaussian Generalized Inverted Wishart Model

B.1 Gaussian Inverted Wishart Model

Let $Y_1, \dots, Y_n$ be the g-dimensional independent random sample and denote $Y' = (Y_1, \dots, Y_n)$. We now consider the distribution of Y as an $n \times g$ random matrix.

Lemma 1. *If*

$$\begin{aligned} &Y \mid \Sigma \sim N(0, I_n \otimes \Sigma) \\ &\Sigma \sim IW(\Psi, \delta) \end{aligned} \tag{33}$$

then

$$\begin{aligned} &\Sigma \mid Y \sim IW(\Psi + Y'Y, \delta + n) \\ &Y \sim t_{n\times g}(0, (\delta - g + 1)^{-1} I_n \otimes \Psi, \delta - g + 1). \end{aligned} \tag{34}$$

Proof. The proof follows stardard arguments. For details, see Anderson (1984)

A similar result for the Gaussian Inverted Wishart model with covariates is presented below.

Lemma 2. *Let $Y : n \times g$, $\boldsymbol{\beta}$, $\boldsymbol{\beta}_0 : l \times g$, $\Sigma > 0, \Psi > 0 : g \times g$, $Z : n \times l$, $F > 0 : l \times l$, and $A > 0 : n \times n$. Given the Gaussian Inverted Wishart model with covariates*

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

we have 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

$$\begin{aligned} &\boldsymbol{\beta} \mid \Sigma, Y \sim N(W\hat{\boldsymbol{\beta}} + (I - W)\boldsymbol{\beta}_0, \tilde{F}^{-1} \otimes \Sigma), \\ &\Sigma \mid Y \sim IW(\Psi + (Y - Z\boldsymbol{\beta}_0)'(A + ZF^{-1}Z')^{-1}(Y - Z\boldsymbol{\beta}_0), \delta + n) \end{aligned}$$

where

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

Proof. By integrating out $\boldsymbol{\beta}$, we have

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

This implies the predictive distribution of Y and the posterior distribution of $\Sigma \mid Y$ using the previous Lemma 2. Finally obtaining the posterior distribution of $\boldsymbol{\beta} \mid \Sigma, Y$ follows a standard line by completing the quadratic function of $\boldsymbol{\beta}$

$$(Y - Z\boldsymbol{\beta})'A^{-1}(Y - Z\boldsymbol{\beta}) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'F(\boldsymbol{\beta} - \boldsymbol{\beta}_0).$$

B.2 Gaussian Generalized Inverted Wishart Model

Let

- $Y_t = \begin{pmatrix} Y_t^{[u]} \\ Y_t^{[g]} \end{pmatrix}$, for $t = 1, \ldots, n$, be the response vector where

  $Y_t^{[u]}$: $u \times 1$, is the response vector at the ungauged sites,

  $Y_t^{[g]}$: $g \times 1$, the response vector at the gauged sites,

- $Y = (Y_1, \ldots Y_n)' = (Y^{[u]}, Y^{[g]})$, be the data matrix where

$$Y^{[u]} = (Y_1^{[u]}, \ldots Y_n^{[u]})', \quad Y^{[g]} = (Y_1^{[g]}, \ldots Y_n^{[g]})'.$$

Consider the following transformation of the partitioned covariance matrix Σ of Y

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

where $\Sigma_{11} : u \times u$, $\Sigma_{21} : g \times u$, $\Sigma_{22} : g \times g$ and

$$\tau = \Sigma_{22}^{-1}\Sigma_{21}, \quad \Gamma = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

The Gaussian and Generalized Inverted Wishart is specified by

$$\begin{cases} Y \mid \Sigma \sim N(0, I_n \otimes \Sigma) \\ \tau \mid \Gamma \sim N(\tau_0, H \otimes \Gamma) \\ \Gamma \sim IW(\Psi_1, \delta_1) \\ \Sigma_{22} \sim IW(\Psi_2, \delta_2). \end{cases} \tag{35}$$

Lemma 3. *The predictive distribution of $(Y^{[u]} \mid Y^{[g]})$ is a matric t-distribution:*

$$(Y^{[u]} \mid Y^{[g]}) \sim t_{n \times u}\left(Y^{[g]}\tau_0, (\delta_1 - u + 1)^{-1}(I_n + Y^{[g]}HY^{[g]'}) \otimes \Psi_1, \delta_1 - u + 1\right). \tag{36}$$

Proof: The conditional distribution of $(Y^{[u]} \mid Y^{[g]}, \Sigma)$, which does not depend on Σ_{22}, is given by

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

The result of the distribution of $(Y^{[u]} \mid Y^{[g]})$ follows by applying Lemma 2 to the Gaussian Inverted Wishart model of $(Y^{[u]}, \tau, \Gamma)$, given $Y^{[g]}$, specified by (35) and (36).

Corollary . *The covariance matrix of $(Y^{[u]}, Y^{[g]})$ can be computed as*

$$\text{var}\left(vec[Y^{[g]}]\right) = I_n \otimes \frac{\Psi_2}{\delta_2 - g - 1},$$

$$\text{cov}\left(vec[Y^{[u]}], vec[Y^{[g]}]\right) = I_n \otimes \frac{\tau_0'\Psi_2}{\delta_2 - g - 1},$$

$$\text{var}\left(vec[Y^{[u]}]\right) = I_n \otimes \left(\left[1 + \frac{\text{tr}(H\Psi_2)}{\delta_2 - g - 1}\right]\frac{\Psi_1}{\delta_1 - u - 1} + \frac{\tau_0'\Psi_2\tau_0}{\delta_2 - g - 1}\right).$$

Consider GIW model with covariates

$$\begin{cases} Y \mid \boldsymbol{\beta}, \Sigma \sim N(Z\boldsymbol{\beta}, A \otimes \Sigma); \\ \boldsymbol{\beta} \mid \Sigma \sim N(\boldsymbol{\beta}_0, F^{-1} \otimes \Sigma); \\ \tau \mid \Gamma \sim N\left(\tau_0, H \otimes \Gamma\right); \\ \Gamma \sim IW(\Psi_1, \delta_1); \\ \Sigma_{22} \sim IW(\Psi_2, \delta_2), \end{cases} \tag{37}$$

where Y and Σ are partitioned as before. Moreover

$$\boldsymbol{\beta} = (\boldsymbol{\beta}^{[u]}, \boldsymbol{\beta}^{[g]}) : (l \times u, l \times g)$$

$$\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_0^{[u]}, \boldsymbol{\beta}_0^{[g]}) : (l \times u, l \times g).$$

Lemma 4. *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

$$\mu^{[u|g]} = Z\boldsymbol{\beta}_0^{[u]} + (Y^{[g]} - Z\boldsymbol{\beta}_0^{[g]})\tau_0,$$

$$\Phi^{[u|g]} = A + ZF^{-1}Z' + (Y^{[g]} - Z\boldsymbol{\beta}_0^{[g]})H(Y^{[g]} - Z\boldsymbol{\beta}_0^{[g]})',$$

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

Proof. Notice that

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

The result follows immediately from the previous Lemma 3.

We now assume there exists an ungauged time period at the gauged sites. The predictive distribution of the unobserved responses is developed below.

Assume

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

where $Y_{(1)}^{[g]} : m \times g,\ m < n$ is not observed. To present the result of the predictive distribution of $Y_{(1)}^{[g]}$ given the observed data $Y_{(2)}^{[g]}$, we introduce the following notation:

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

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

Moreover,

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

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

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

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

Lemma 5. *The predictive distribution of $Y^{[g]}_{(1)}$ given data $Y^{[g]}_{(2)}$ is*

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

Proof. The GIW model (37) implies the GIW sub-model

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

Therefore,

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

Using Proposition 3, we have

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

This proves the Lemma.

C Proofs

The proof of Theorem 1. Part (i) is an immediate result of Lemma 5, Part (iii), an immediate result of Lemma 4, while Part (ii) is obtained by first applying Lemma 4 to $Y^{[g_j]}$ given $Y^{[g_{j+1},\dots,g_k]}$, then applying Lemma 5 to $Y^{[g_j^1]}$ given $Y^{[g_j^2]}$ and $Y^{[g_{j+1},\dots,g_k]}$.

The proof of Corollary 1. The proof is straightforward if we notice that $\mu^{[j]}_{(2)}$ and $A^{[j]}_{22}$ do not depend on unobserved responses $Y^{[g^1_{j+1}]}, \dots, Y^{[g^1_k]}$.

The proof of Theorem 2. (i) The model (1) implies

$$\begin{cases} \boldsymbol{\beta}^{[g_k]} \mid \Sigma_{kk}, \boldsymbol{\beta}_0, F \sim N(\boldsymbol{\beta}^{[k]}_0, F^{-1} \otimes \Sigma_{kk}); \\ Y^{[g^2_k]} \mid \boldsymbol{\beta}^{[g_k]}, \Sigma_{kk}, \mathcal{H} \sim N(Z_{(k)}\boldsymbol{\beta}^{[g_k]}, I_{n-m_k} \otimes \Sigma_{kk}), \end{cases} \tag{38}$$

which gives the posterior distribution of $(\boldsymbol{\beta}^{[g_k]} \mid D, \Sigma_{kk}, \mathcal{H})$, as in (9), by means of Lemma 2.

Similarly, model (1) implies

$$\left\{ \begin{array}{l} \boldsymbol{\beta}^{[g_j]} \mid \boldsymbol{\beta}^{[g_{j+1},\ldots,g_k]}, \tau_j, \Gamma_j, \mathcal{H} \sim N\left(\boldsymbol{\beta}_0^{[g_j]} + (\boldsymbol{\beta}^{[g_{j+1},\ldots,g_k]} - \boldsymbol{\beta}_0^{[g_{j+1},\ldots,g_k]})\tau_j, F^{-1}\otimes\Gamma_j\right), \\ \\ Y^{[g_j^2]} \mid Y_{(j)}^{[g_{j+1},\ldots,g_k]}, \boldsymbol{\beta}^{[g_j,g_{j+1},\ldots,g_k]}, \tau_j, \Gamma_j, \mathcal{H} \sim N(Z_{(j)}\boldsymbol{\beta}^{[g_j]} + \tilde{\epsilon}_{(j)}^{[g_{j+1},\ldots,g_k]}\tau_j, F^{-1}\otimes\Gamma_j), \end{array} \right. \tag{39}$$

where

$$\tilde{\epsilon}_{(j)}^{[g_{j+1},\ldots,g_k]} = Y_{(j)}^{[g_{j+1},\ldots,g_k]} - Z_{(j)}\boldsymbol{\beta}_0^{[g_{j+1},\ldots,g_k]}.$$

Applying Lemma 2 again gives the posterior of $(\boldsymbol{\beta}^{[g_j]} \mid D, \boldsymbol{\beta}^{[g_{j+1},\ldots,g_k]}, \tau_j, \Gamma_j, \mathcal{H})$ as in (9).

(ii) Combining (38), (39) and (3) and using the result of Lemma 2, one can the result similarly.

The proof of Corollary 2. Straightforward.

D Calculation of Posterior Expectations

Lemma 6.: *The posterior means of $\boldsymbol{\beta}$ given data D and hyperparameter set $\mathcal{H}$ are given below:*

$$E(\boldsymbol{\beta}^{[g_k]} \mid D, \mathcal{H}) = \tilde{\boldsymbol{\beta}}^{[g_k^2]},$$

$$E(\boldsymbol{\beta}^{[g_j]} \mid D, \mathcal{H}) = \tilde{\boldsymbol{\beta}}^{[g_j^2]} + \left[E(\boldsymbol{\beta}^{[g_{j+1},\ldots,g_k]} \mid D, \mathcal{H}) - \tilde{\boldsymbol{\beta}}^{[g_{j+1},\ldots,g_k]}\right]\tilde{\tau}_j,$$

where

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

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

Proof. Straightforward by using Theorem 2.

Lemma 7.: *The posterior mean of $\Sigma^{-1} = \Sigma_{(11)}^{-1}$ given data D and hyperparameter set $\mathcal{H}$ is computed recursively as following:*

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

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

$$E\left[\Sigma_{(kk)}^{-1} \mid D, \mathcal{H}\right] = \tilde{\delta}_k\tilde{\Psi}_k^{-1}.$$

Proof. Notice that

$$\Sigma_{(jj)}^{-1} = \begin{pmatrix} \Gamma_j^{-1} & -\Gamma_j^{-1}\tau_j' \\ \\ -\tau_j\Gamma_j^{-1} & \tau_j\Gamma_j^{-1}\tau_j' + \Sigma_{(j+1,j+1)}^{-1} \end{pmatrix}, \text{ for } j = 1, \ldots, k-1.$$

Using Theorem 2,

$$E\left[\Gamma_j^{-1} \mid D, \mathcal{H}\right] = \tilde{\delta}_j\tilde{\Psi}_j^{-1},$$

$$E\left[\Gamma_j^{-1}\tau_j' \mid D, \mathcal{H}\right] = \tilde{\delta}_j\tilde{\Psi}_j^{-1}\tilde{\tau}_j',$$

$$E\left[\tau_j\Gamma_j^{-1}\tau_j' \mid D, \mathcal{H}\right] = \tilde{\delta}_j\tilde{\tau}_j\tilde{\Psi}_j^{-1}\tilde{\tau}_j' + g_j\tilde{H}_j,$$

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

$$E\left[\Sigma_{(kk)}^{-1} \mid D, \mathcal{H}\right] = E\left[\Sigma_{kk}^{-1} \mid D, \mathcal{H}\right] = \tilde{\delta}_k \tilde{\Psi}_k^{-1}.$$

This proves the Lemma.

Lemma 8.: *Let η denote the digamma function. Then we have*

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

and

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

Lemma 9.: *The posterior expectation $E\left\{\boldsymbol{\beta}\Sigma^{-1} \mid D, \mathcal{H}\right\}$ is obtained as following:*

$$E\left\{\boldsymbol{\beta}\Sigma^{-1} \mid D, \mathcal{H}\right\} = E\left\{\boldsymbol{\beta}^{[g_1,\dots,g_k]}\Sigma_{(11)}^{-1} \mid D, \mathcal{H}\right\},$$

$$E\left\{\boldsymbol{\beta}^{[g_j,\dots,g_k]}\Sigma_{(jj)}^{-1} \mid D, \mathcal{H}\right\} = \Big(\tilde{\delta}_j(\tilde{\boldsymbol{\beta}}^{[g_j^2]} - \tilde{\boldsymbol{\beta}}^{[g_{j+1},\dots,g_k]}\tilde{\tau}_j)\tilde{\Psi}_j^{-1},$$

$$E\left\{\boldsymbol{\beta}^{[g_{j+1},\dots,g_k]}\Sigma_{(j+1,j+1)}^{-1} \mid D, \mathcal{H}\right\} - \tilde{\delta}_j(\tilde{\boldsymbol{\beta}}^{[g_j^2]} - \tilde{\boldsymbol{\beta}}^{[g_{j+1},\dots,g_k]}\tilde{\tau}_j)\tilde{\Psi}_j^{-1}\tilde{\tau}_j' + g_j\tilde{\boldsymbol{\beta}}^{[g_{j+1},\dots,g_k]}\tilde{H}_j\Big),$$

and at the last

$$E\left\{\boldsymbol{\beta}^{[g_k]}\Sigma_{kk}^{-1} \mid D, \mathcal{H}\right\} = \tilde{\delta}_k\tilde{\boldsymbol{\beta}}^{[g_k]}\tilde{\Psi}_k^{-1}.$$

Proof. We can write

$$\boldsymbol{\beta}^{[g_j,\dots,g_k]}\Sigma_{(jj)}^{-1}$$

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

The rest is straightforward.

Lemma 10.: *The posterior expectation $E\left\{\boldsymbol{\beta}\Sigma^{-1}\boldsymbol{\beta}' \mid D, \mathcal{H}\right\}$ is obtained as following:*

$$E\left\{\boldsymbol{\beta}\Sigma^{-1}\boldsymbol{\beta}' \mid D, \mathcal{H}\right\} = \tilde{\delta}_k\tilde{\boldsymbol{\beta}}^{[g_k^2]}\tilde{\Psi}_k^{-1}\tilde{\boldsymbol{\beta}}^{[g_k^2]'}$$

$$+ \textstyle\sum_{j=1}^{k-1} \tilde{\delta}_j \left[\tilde{\boldsymbol{\beta}}^{[g_j^2]} - \tilde{\boldsymbol{\beta}}^{[g_{j+1},\dots,g_k]})\tilde{\tau}_j\right] \tilde{\Psi}_j^{-1} \left[\tilde{\boldsymbol{\beta}}^{[g_j^2]} - \tilde{\boldsymbol{\beta}}^{[g_{j+1},\dots,g_k]})\tilde{\tau}_j\right]'$$

$$+ \textstyle\sum_{j=1}^{k-1} g_j\tilde{\boldsymbol{\beta}}^{[g_{j+1},\dots,g_k]}\tilde{H}_j\tilde{\boldsymbol{\beta}}^{[g_{j+1},\dots,g_k]'} + \sum_{j=1}^{k} g_j\tilde{F}_j^{-1}$$

Proof. The proof is obtained by observing that

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

References

[1] Anderson, TW (1984) An Introduction To Multivariate Statistical Analysis. (2nd Ed.) Wiley.

[2] Barnett, V and O'Hagan, A (1997) Setting environmental standards: The statistical approach to handling uncertainty and variation. London: Chapman and Hall.

[3] Brauer, M, Duddek, C, Hornby, S, Le, ND, Stevens, J, Sun, W, White, R, Zhang, H and Zidek, JV (1996). An assessment of pNEM: a computer package for assessing population level exposures to air pollution. TR #162, Department of Statistics, University of British Columbia.

[4] Brown, PJ, Le, ND and Zidek, JV (1994a). Multivariate Spatial Interpolation And Exposure to Air Pollutants *Can Jour Statist* 22:489-510.

[5] Brown, PJ, Le, ND and Zidek, JV (1994b). Inference for a covariance matrix. In Smith, AFM and Freeman, PR eds. *Aspects of Uncertainty: A Tribute to D.V. Lindley.* Wiley.

[6] Burnett, R and Krewski, D (1994). Air pollution effects on hospital admission rates: a random effects modelling approach, *Can Jour Statist* 22:41-458.

[7] Caselton, WF and Zidek, JV (1984). Optimal Monitoring Network Designs. *Statist Prob Letters* 2:223-227.

[8] Caselton, WF, Kan, L and Zidek, JV (1992). Quality Data Networks that Minimize Entropy. *Statistics in the Environmental and Earth Sciences.* Eds. P. Guttorp and Walden. Griffin, London.

[9] Chen, CF (1979). Bayesian Inference for a Normal Dispersion Matrix and its Application to Stochastic Multiple Regression Analysis. *Journal Royal Statistical Society, Series B* 41:235-48.

[10] Cressie, NC (1991). *Statistics For Spatial Data.* Wiley.

[11] Davidian, M and Giltinan, DM (1995). *Nonlinear models for repeated measurement data.* London: Chapman and Hall.

[12] Dempster, AP, Laird, NM and Rubin, DB (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal Royal Statistical Society, Series B.* 39:1-38.

[13] Duddek, C, Le, ND, Zidek, JV and Burnett, RT (1995). Multivariate Imputation in Cross Sectional Analysis of Health Effects Associated with Air Pollution (with discussions). Journal of Environmental and Ecological Statistics. 2:191-212.

[14] Fuller, W (1987). *Measurement Error Models.* New York:Wiley.

[15] Guttorp, P, Le, ND, Sampson, PD and Zidek, JV (1993). Using Entropy in the Redesign of an Environmental Monitoring Network. *Multivariate Environmental Statistics*; edited by GP Patil and CR Rao. North Holland/Elsevier Science, NY.

[16] Ko, C-W, Lee, J and Queyranne, M (1995). An Exact Algorithm for Maximum Entropy sampling. *Operations Res* 43:684-691.

[17] Le, ND and Zidek, JV (1992). Interpolation with uncertain spatial covariances: a Bayesian alternative to kriging. *Journal of Multivariate Analysis* 43:351-74.

[18] Le, ND and Zidek, JV (1994). Network Designs for Monitoring Multivariate Random Spatial Fields. In *Recent Adv in Stat and Prob*; edited by JP Vilaplana and ML Puri 191-206.

[19] Le, ND, Sun, W and Zidek, JV (1997). Bayesian Multivariate Spatial Interpolation with Data Missing-by-Design. *Jour Roy Statist Soc, Series B* 59:501-510.

[20] Le, ND, Sun, L and Zidek, JV (1998). A Note On the Existence of Maximum Likelihood Estimates for Gaussian-Inverted Wishart Models. *Stat & Prob Letters* 40:133-137.

[21] Pierce, DA, Stram, DO, Vaeth, M and Schafer, DW (1992), The errors-in-variables problem: considerations provided by radiation dose-response analyses of the a-bomb survivor data. *Jour Amer Statist Assoc*:87:351-359.

[22] Sampson, PD and Guttorp, P (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association* 87:108-19.

[23] Sun, W (1994). Bayesian multivariate interpolation with missing data and its applications. Ph.D. Thesis, University of British Columbiar, Vancouver, BC, Canada.

[24] Sun, W (1995). Comparison of a co-Kriging method with a Bayesian alternative. TR No 147, Department of Statistics, University of British Columbia.

[25] Trees, PR, MacCallum, C, Blackwell, BR, Caldwell, DSC, Joe, H, McIntyre, BW, Biggar, BA, Watkinson, AP, Zidek, JV (1986). High efficiency capture of SO_2, NO_x and heavy metals in combustion flue gases for control of acid rain pollutants. A proposal submitted to the AEERL/EPA in support of a request for funding.

[26] Zeger, SL, Liang, KY and Albert, PS (1988). Models for longitudinal data; a generalised estimating equation approach. *Biometrics* 44:1049-1060.

[27] Zidek, JV, Le, ND (1999) ENVIRONMENTAL STATISTICS: ANALYSING DATA FOR ENVIRONMENTAL POLICY - Measuring and Modelling Pollution for Risk Analysis. *Novartis Foundation Symposium* 220:105-121. Wiley.

[28] Zidek, JV, Wong, H, Le, ND and Burnett, RT (1996). Causality, measurement error and multicollinearity in epidemiology. *Environmetrics* 7:441-451.

[29] Zidek, JV, Le, ND, Wong, H and Burnett, RT (1998a). Including structural measurement errors in the nonlinear regression analysis of clustered data. *Can Jour Stat.* To appear.

[30] Zidek, JV, White, R, Le, ND, Sun, W and Burnett, RT(1998b). Imputing Unmeasured Explanatory Variables in Environmental Epidemiology With Application To Health Impact Analysis of Air Pollution. *J Environmental and Ecological Statist.*. To Appear.

[31] Zidek, JV (1997). Interpolating air pollution for health impact assessment. *Statistics for the Environment 3: Pollution Assessment and Control.* Edited by V Barnett and K Feridun Turkman. New York: Wiley 251-268.

[32] Zidek, JV, Sun, W and Le, ND. (1998). Designing and Integrating Composite Networks For Monitoring Multivariate Gaussian Pollution Fields. *Journal Royal Statistical Society.* Submitted.

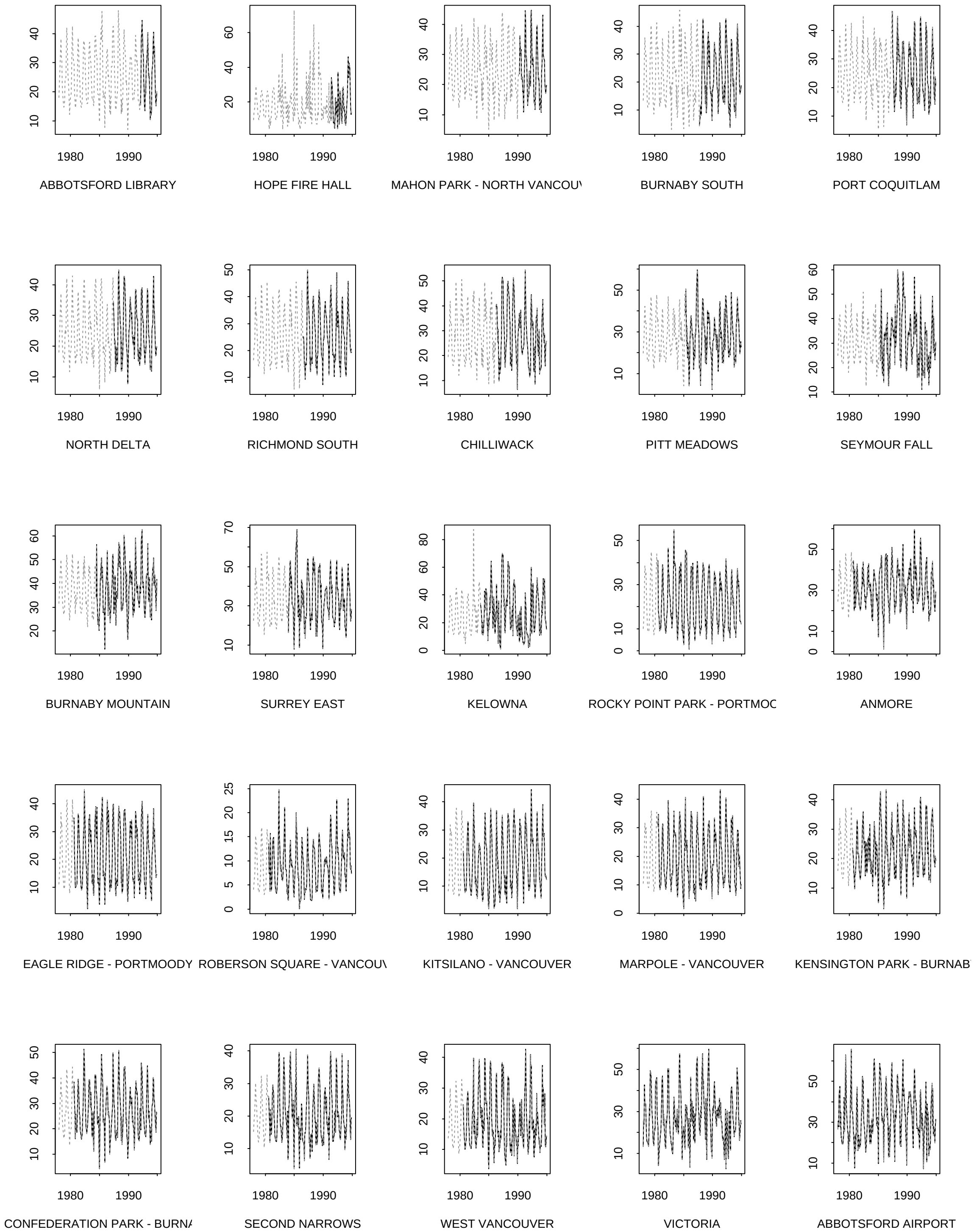

Figure 1: Backcasting

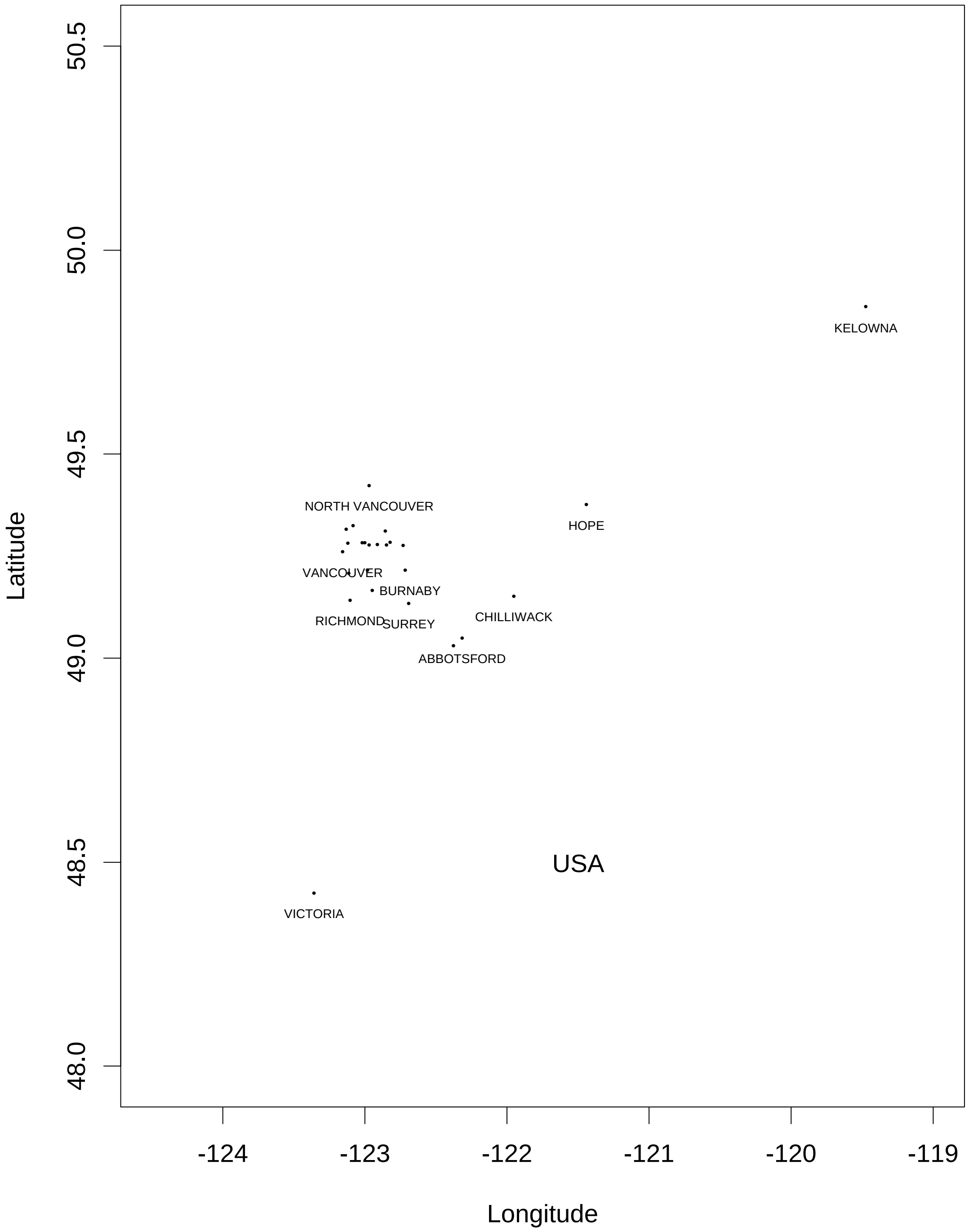

Figure 2: Ozone Monitor Sites in BC, 1978-1994

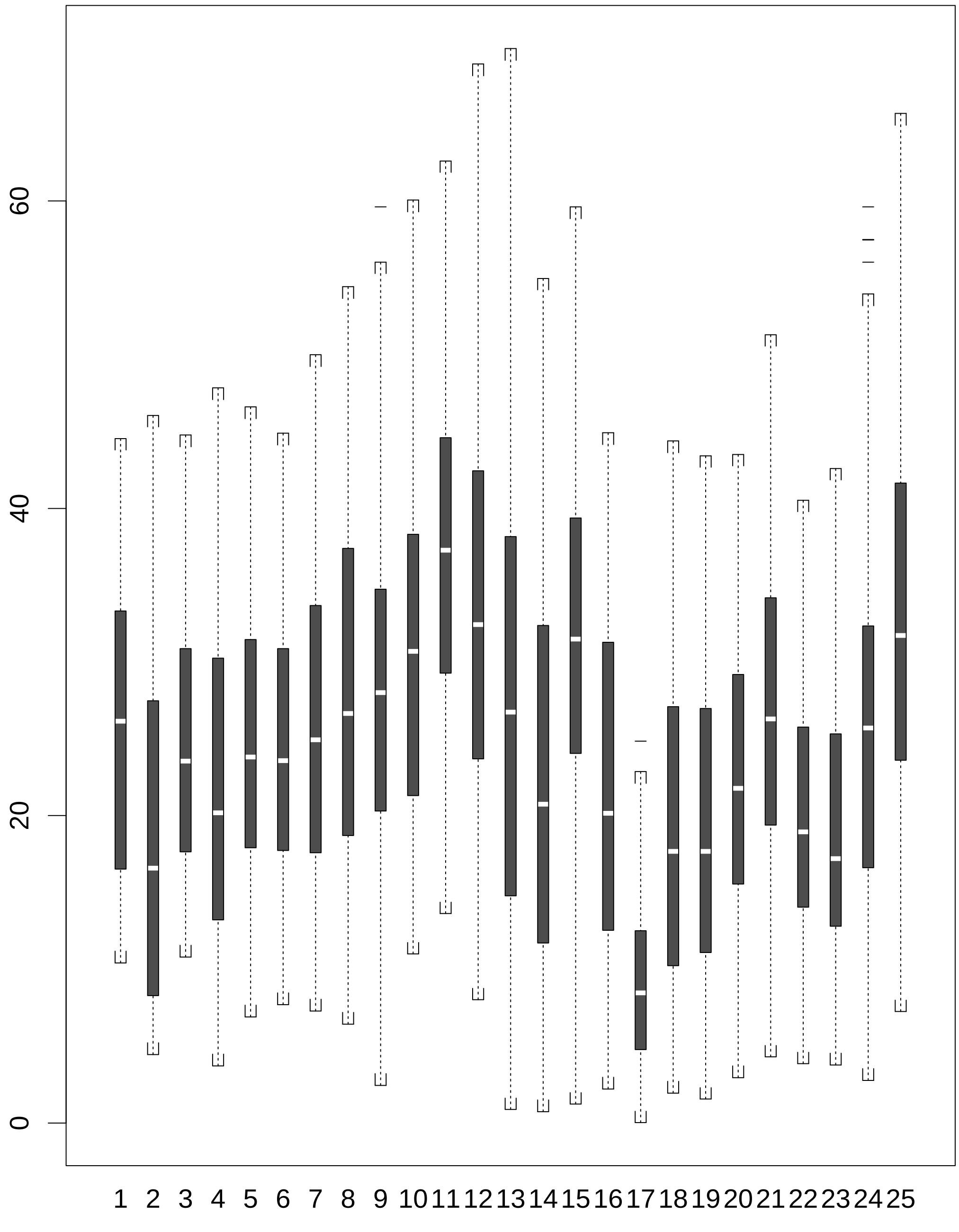

Figure 3: Boxplot of Ozone Levels at 25 Monitor Sites in BC

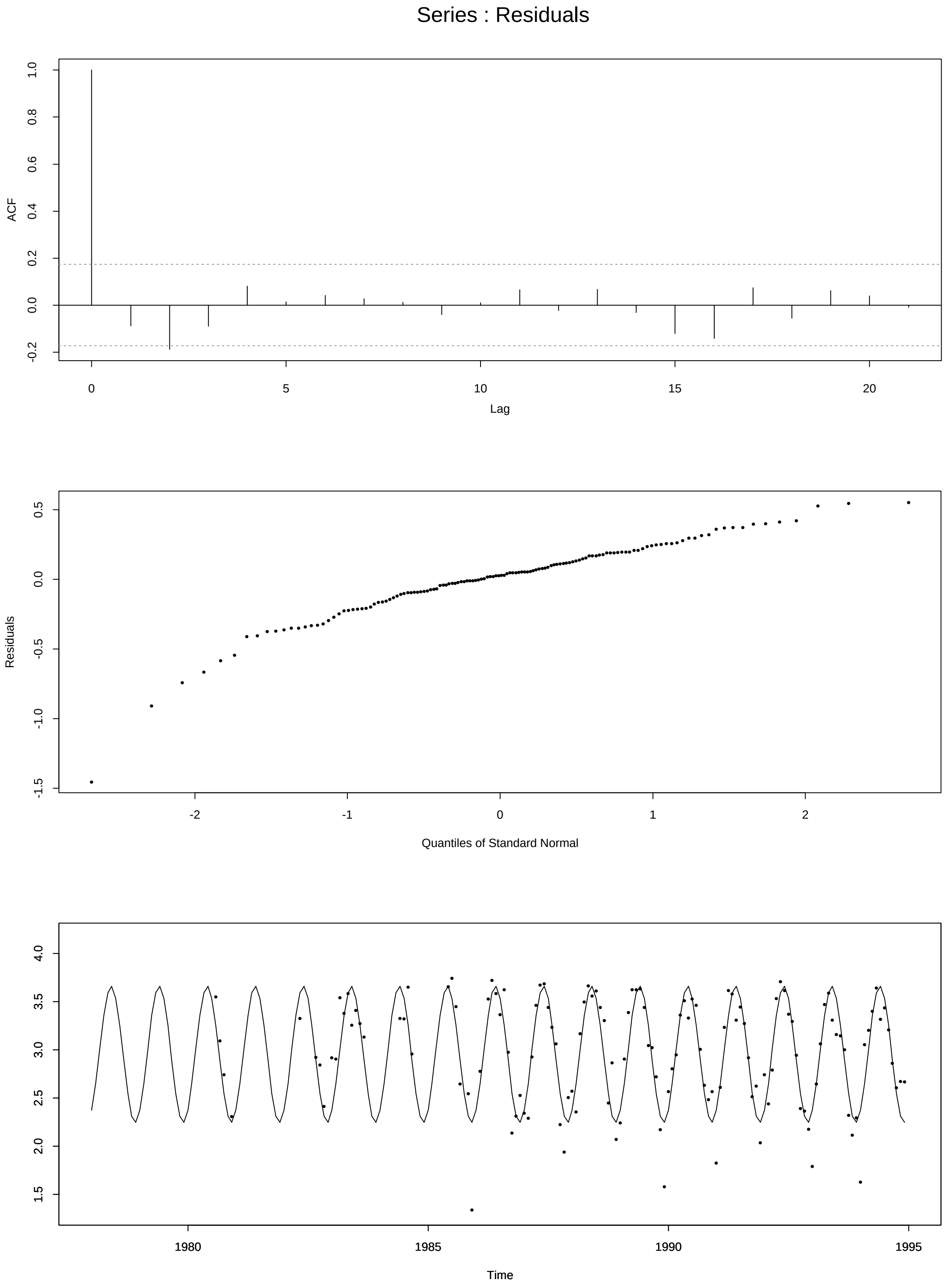

Figure 4: Trend modelling: Upper: autocorrelation function; Middle: standard normal plot; Lower: fitted trend and observations

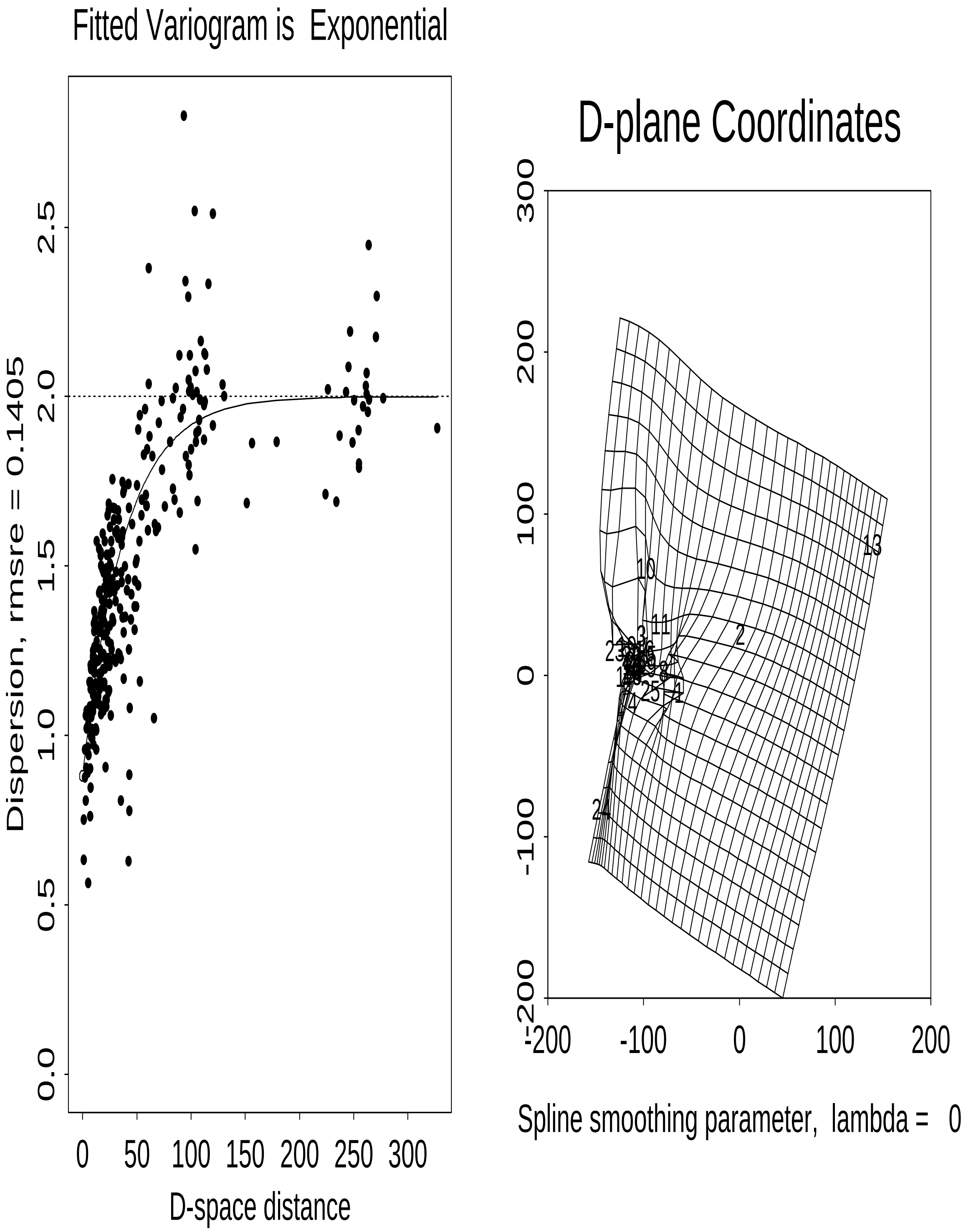

Figure 5: Variogram Fitts and D-plane Smoothing ($\lambda = 0$)

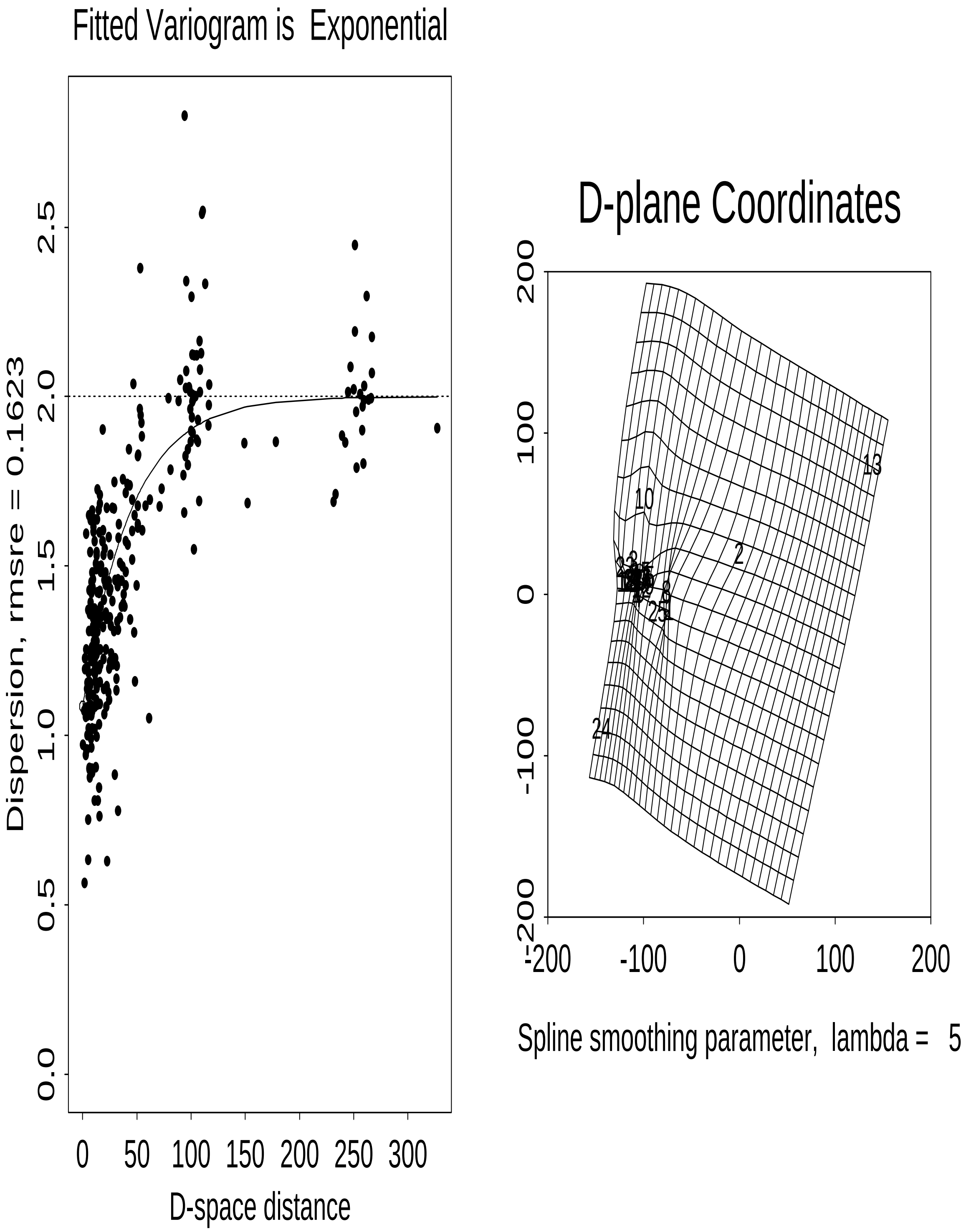

Figure 6: Variogram Fitts and D-plane Smoothing ($\lambda = 5$)

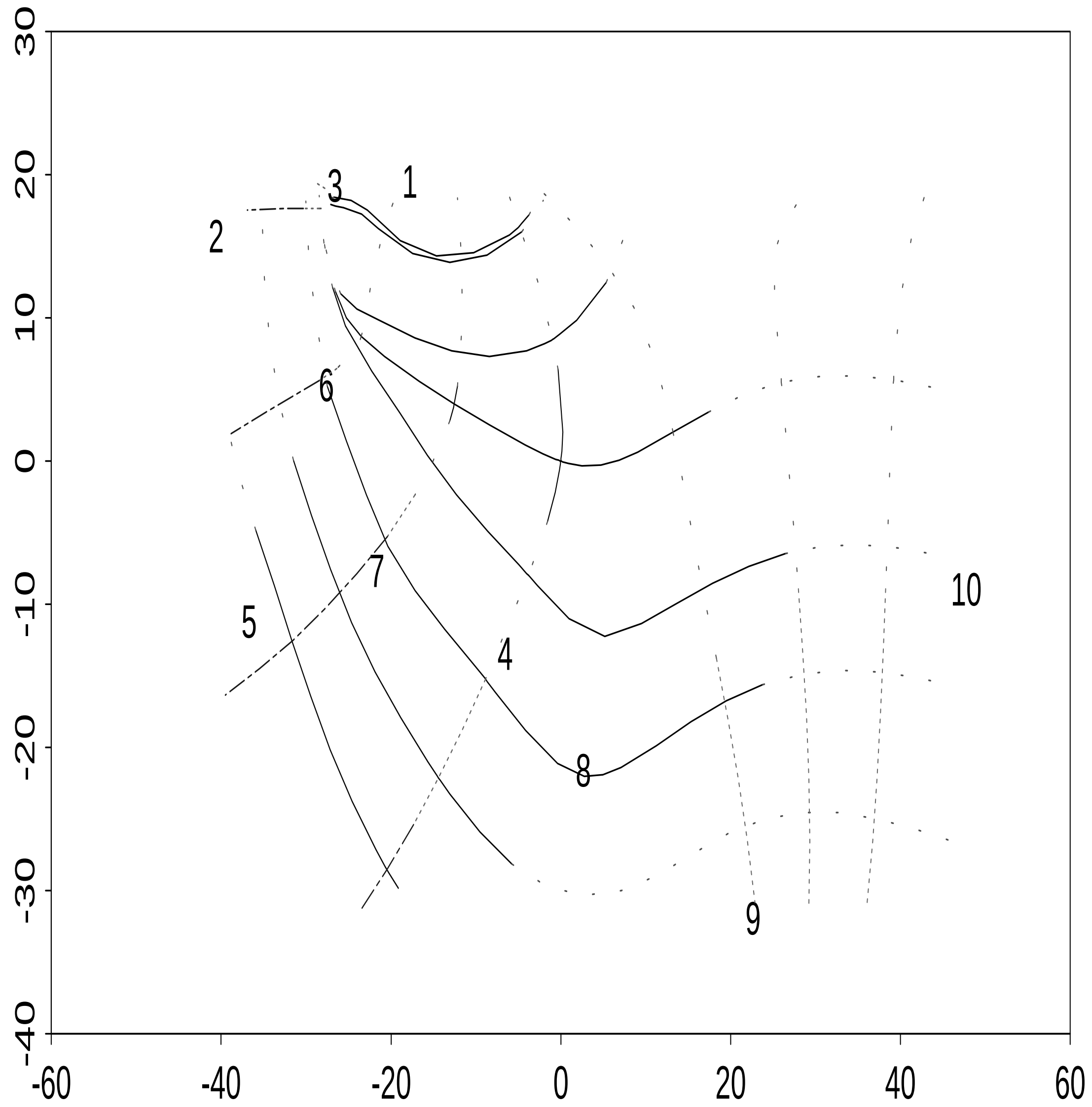

Figure 7: G-space transformation - biorthogonal grid

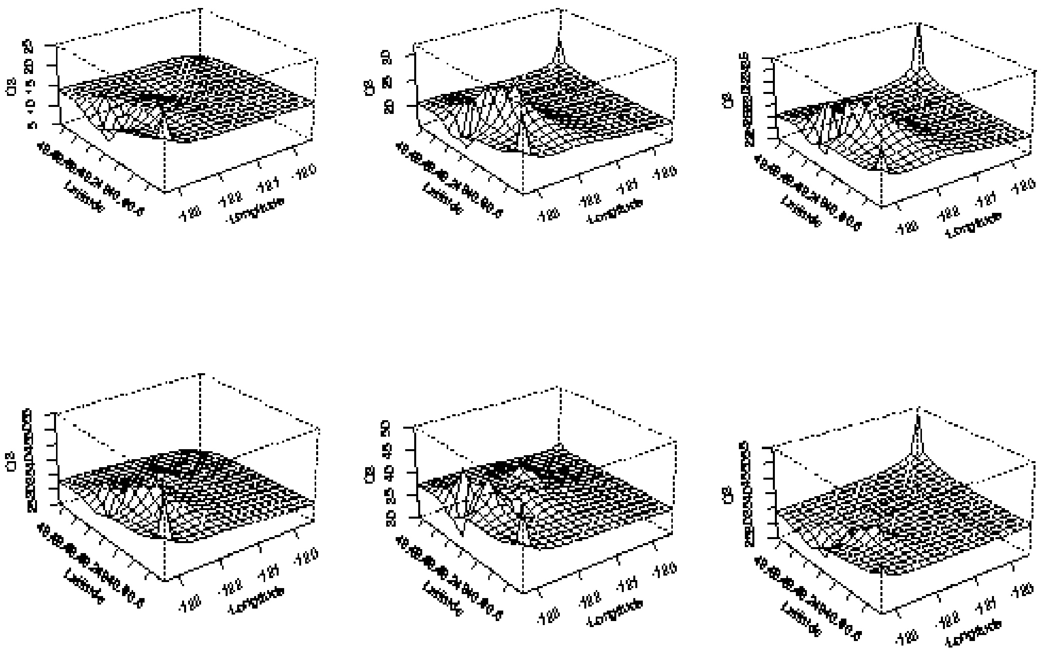

Figure 8: Interpolation of Ozone Levels, January-June 1994

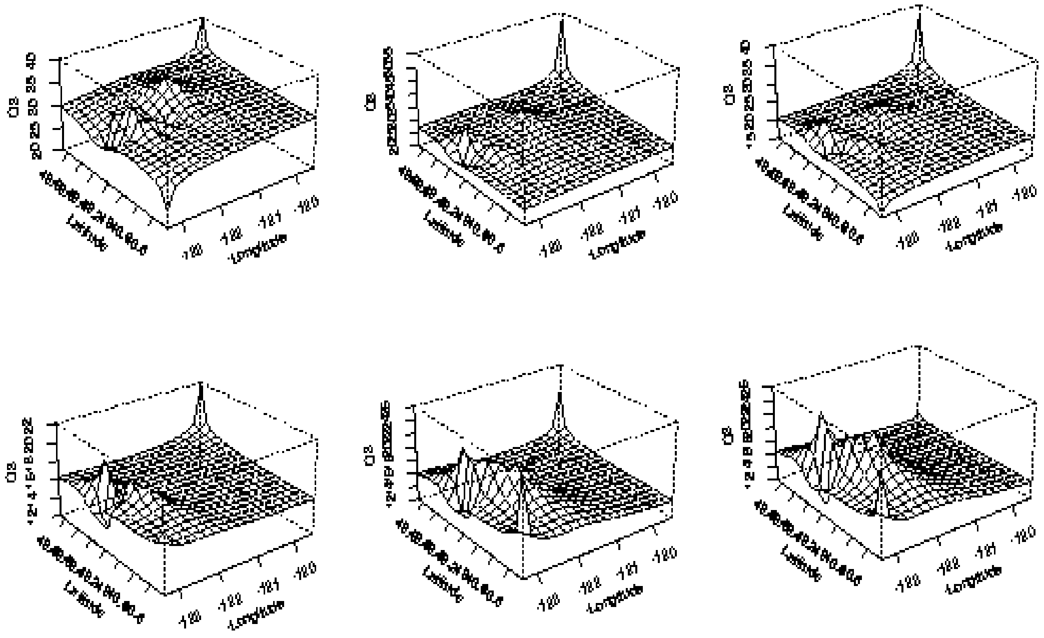

Figure 9: Interpolation of Ozone Levels, July-December 1994

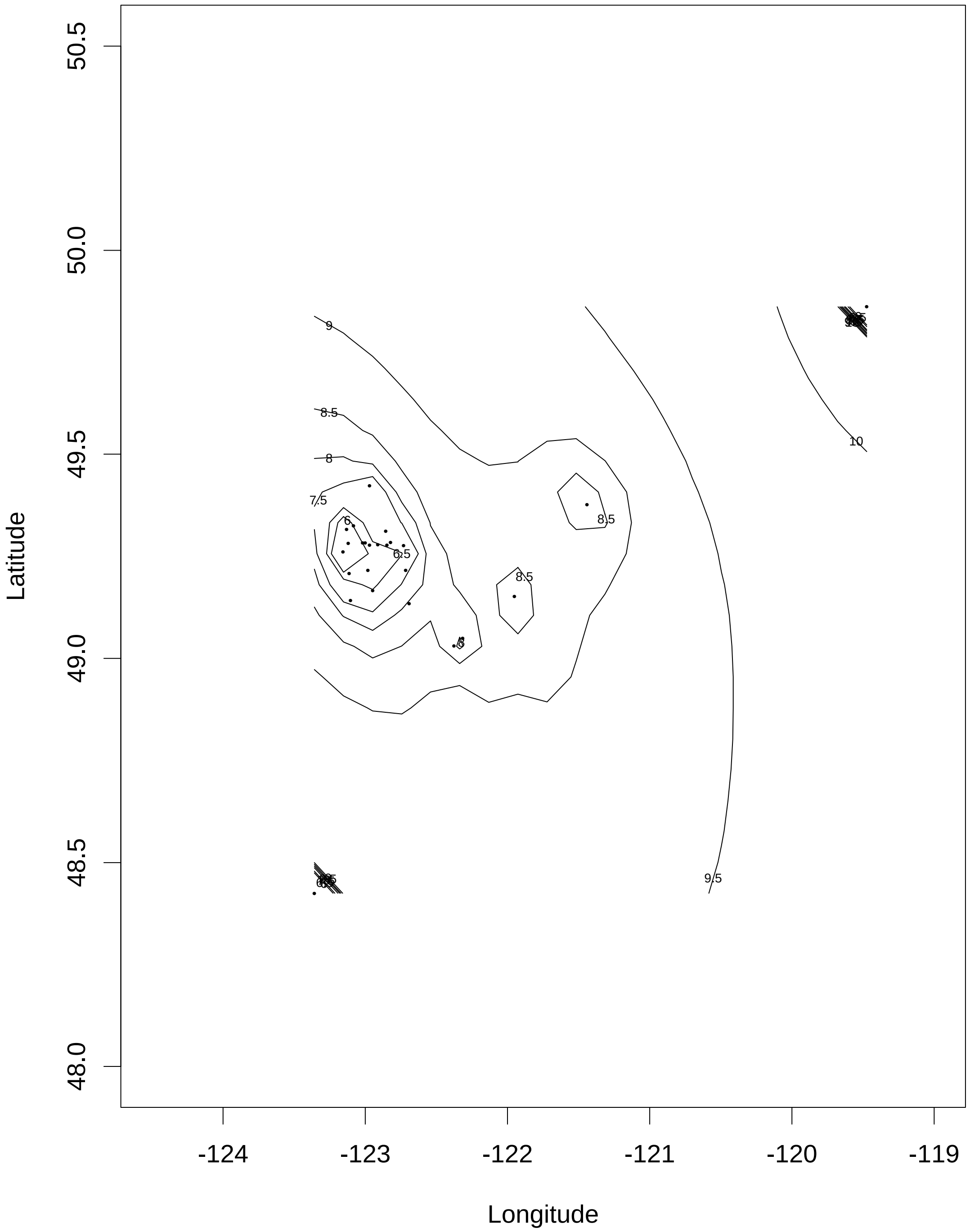

Figure 10: Contour Plot of The SDs of The Interpolations in June, 1994