A BAYESIAN APPROACH TO BACKCASTING AND SPATIALLY PREDICTING UNMEASURED MULTIVARIATE RANDOM SPACE-TIME FIELDS WITH APPLICATION TO $PM_{2.5}$

by

B.M. Golam Kibria, Li Sun, James V. Zidek,
and Nhu D. Le

Technical Report #193

July 2000

A Bayesian Approach to Backcasting and Spatially Predicting Unmeasured Multivariate Random Space-Time Fields With Application to $PM_{2.5}$

B. M. Golam Kibria, Li Sun, James V. Zidek
Department of Statistics
University of British Columbia
Vancouver, BC, V6T 1Z2
and
Nhu D. Le
University of British Columbia, and
British Columbia Cancer Agency
Vancouver, BC

Abstract

This paper presents a multivariate spatial prediction methodology in a Bayesian framework. The method is especially suited for use in environmetrics, where vector-valued responses are observed at a small set of ambient monitoring stations ('gauged sites') at successive time points. However the stations have varying start-up times so that the data has a 'staircase' pattern ('monotone' pattern in the terminology of Rubin and Shaffer 1990). The lowest step corresponds to the newest station in the monitoring network.

We base our approach on a hierarchical Bayes prior involving a Gaussian generalized inverted Wishart (GIW) model. For given hyperparameters, we derive the predictive distribution for currently gauged sites at times prior to their start-up when no measurements were taken.

The resulting predictive distribution is a matrix T distribution with appropriate covariance parameters and degrees of freedom. We estimate the hyperparameters using the method of moment (MOM) as an easy-to-implement alternative to the more complex, Expectation and Maximization(EM) algorithm. The MOM in particular gives exact parameter estimates and involves less cumbersome calculations than the EM algorithm.

Finally we obtain the predictive distribution for unmeasured responses at 'ungauged' sites. The results obtained in this paper allow us to pool the data from sites that measure different pollutants and also to treat cases where the observed data monitoring station have a monotonic 'staircase' structure.

We demonstrate the use of this methodology by mapping $PM_{2.5}$ fields for Philadelphia during the period May 1992 to September 1993. Large amounts of data missing by design make this application particularly challenging. However give empirical evidence that the method performs well.

Key words and Phrases: Bayes; Hierarchical; Inverted Wishart Distribution; Matrix T- Distribution; Method of Moment; Posterior Distribution; Predictive Distribution; Spatial Interpolation.

1 Introduction

A strong association between particulate pollution fields and a variety of negative human health outcomes has been demonstrated through a series of epidemiological studies (*c.f.* Bates et al 1990, Burnett et al 1994, Dockery et al 1992 and 1993, Pope et al 1995, Schwarz et al 1993, Ostro et al 1991 and Roemer et al 1993). Interest has concentrated particularly on two particles sizes, $PM_{2.5}$ and PM_{10}. The smaller of these two (and even finer particles) is commonly thought to be a greater risk factor. However Smith et al (2000) recently published evidence to the contrary and provide a source of references to on this issue.

Critics have argued that measurement error compromises the quality of these epidemiological studies (*c.f.* Levy et al 2000). In particular, epidemiological studies have often relied on measurements obtained at ambient monitoring sites even though these may be situated a long way away from the subjects in the study. Thus they do not forecast well the true personal exposure sustained by these subjects. This deficiency and the criticism that can ensure has led to the recognition of the need to spatially predict *i.e.* interpolate random pollution fields with a view to getting improved estimates of true exposure.

Such studies may include health outcomes over a long period of time, for example 6 years in the case of Zidek et al (1997). Over that such long periods not all currently gauged stations will necessarily have been in operation. In this case, investigators will need to 'impute' or 'backcast' in the terminology of this paper the unmeasured values at these sites in addition to predicting unmeasured values at the sites, such as the centroids of census tracts, where measurements were never made. For chronic diseases with long latency such as cancer where cumulative exposure is likely more relevant, the estimates for concentration levels are particularly important. Imputed levels like these have been used in a number of studies (c.f. Duddek et al 1995, Zidek et al 1997).

In recent years, a Bayesian methodology for both temporal and spatial interpolation has been developed starting with that of Le and Zidek (1992) as an alternative to the well-known method of Kriging (c.f Cressie 1991). The method has further been developed by Brown et al (1994a) and Le et al (1997) to deal with the multivariate setting where possibly not all monitored sites measure the same set of pollutants. 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. The problem of finding Bayesian spatial prediction has been considered by various researchers; Kitanidis (1986), Handcock et al (1993), Cui et al (1995), De Oliveira et al (1997), Gaudard et al (1999) and recently Kibria (2000) to mention a few. The specific problem of estimating the covariance fields for a non-stationary random field has been considered by Sampson and Guttorp (1992) and further in Meiring et al (1997). This problem has received a lot of considerable recent attention notably from Smith (1996), Samian et al (2000), as well as Schmitt and O'hagan (2000).

In the paper we extend the univariate theory of Le et al (1999) to the multivariate case and thereby obtain an empirical hierarchical Bayesian method for temporal and spatial interpolation using all available data. We base the method on the Gaussian and Generalized Inverted Wishart (GIW) distributions. Specifically, we assume the responses follow a Gaussian distribution and the corresponding covariance follows a Generalized Inverted Wishart (GIW) prior distribution. We adopt a Kronecker product in the GIW model for the covariance matrix. This parametrization reduces the number of parameters and more importantly allows us to estimate the parameters in ungauged sites, *i.e.* census tracts in the application treated in this paper where we do not have any observed data. Although the extension is somewhat formalistic, it has advantages over Le et al (1999). First, this paper will help to simplify the burden of finding estimates of the potentially very large dimensional hyperparameters. Secondly, it will enable estimation of the hyperparameters using the method of Sampson and Guttorp (1992; hereafter SG method or algorithm) involving the predictive distributions for ungauged sites. We develop a method of moments approach as an alternative to the much more computationally intensive EM algorithm, which has provided in the Appendix. However the computer (computational) program for the EM algorithm yet to be developed, which is under the future investigation. It is noted that we assume there are no randomly missing data and return to this issue in the application.

We structure this paper as follows. The main theoretical results about the predictive distribution for gauged sites, those at which measurement are taken are described in Section 2. Parameter estimation using the method of moment is discussed in Section 3. The predictive distribution for the unmeasured responses at ungauged sites and the related parameters estimated by the SG method follows in Section 4. As an application, Philadelphia data are been considered in section 5. Finally, a summary and concluding remarks appear in Section 6.

2 Main Results

2.1 Notation

For notational convenience, let:

$$
\begin{aligned}
n &= \text{number of time points (eg. number of months);} \\
u &= \text{number of locations with no monitors - called ungauged sites;} \\
g &= \text{number of locations with monitors- called gauged sites;} \\
&= g_1 + g_2 + \ldots + g_k; \\
p &= \text{number of responses at both ungauged and gauged sites.}
\end{aligned}
$$

The g gauged sites are organized into k blocks of stations where the g_j ($j = 1, 2, \ldots, k$) stations in the jth block have the same number of time points m_j at which no measurements are taken by design. These blocks are numbered so that the observed measurements have a "rising a staircase" structure, that is,

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

Denote

- the response variables partitioned for the gauged and ungauged sites by

$$Y = \left[Y^{[u]}, Y^{[g]}\right],$$

where

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

an $n \times gp$ matrix, denotes the unobserved and observed responses at gauged sites, and

  - $Y^{[u]}$, an $n \times up$ matrix, denotes the unobserved responses at ungauged sites;
  - $Y^{[g_j^{[1]}]}$, an $m_j \times g_j p$ matrix, denotes the unobserved responses at the g_j gauged sites for the m_j time points;

- $Y^{[g_j^o]}$, an $(n-m_j)\times g_jp$ matrix, denotes the observed responses at the g_j gauged sites for the $(n-m_j)$ time points;

- observed measurements at the gauged sites by

$$D=\left\{Y_2^{[g_1 g_2\cdots g_k]}\right\}=\left\{Y^{[g_1^o]},Y^{[g_2^o]},\cdots,Y^{[g_k^o]}\right\};$$

- unobserved responses by

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

- unobserved responses in gauged sites in blocks j to k by

$$Y^{[g_j^u\cdots g_k^u]}=\left\{Y^{[g_j^u]},\ldots,Y^{[g_k^u]}\right\};$$

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

$$Y^{[g_j\cdots g_k]}=\begin{pmatrix} Y_1^{[g_j\cdots g_k]} \\ Y_2^{[g_j\cdots g_k]}\end{pmatrix}=\left[\begin{pmatrix} Y^{[g_j^u]} \\ Y^{[g_j^o]}\end{pmatrix},\cdots,\begin{pmatrix} Y^{[g_k^u]} \\ Y^{[g_k^o]}\end{pmatrix}\right];$$

- the responses from all gauged sites by

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

The covariance matrix Σ of dimension $(u+g)p\times(u+g)p$ over gauged and ungauged sites are partitioned conformably as

$$\Sigma=\begin{pmatrix} \Sigma^{[u]} & \Sigma^{[ug]} \\ \Sigma^{[gu]} & \Sigma^{[g]}\end{pmatrix},$$

where $\Sigma^{[u]}$ is an $up\times up$ matrix. The covariance matrix $\Sigma^{[g]}$ of dimension $gp\times gp$ on the gauged sites is further partioned by blocks as

$$\Sigma^{[g_j\cdots g_k]}=\begin{pmatrix} \Sigma^{[g_j]} & \ldots & \Sigma^{[g_j g_k]} \\ \ldots & \ldots & \ldots \\ \Sigma^{[g_k g_j]} & \ldots & \Sigma^{[g_k]}\end{pmatrix} \quad\text{and}\quad \Sigma^{[g_1\cdots g_k]}=\begin{pmatrix} \Sigma^{[g_1]} & \ldots & \Sigma^{[g_1 g_k]} \\ \ldots & \ldots & \ldots \\ \Sigma^{[g_k g_1]} & \ldots & \Sigma^{[g_k]}\end{pmatrix}.$$

The following 1-1 transformation (Barlett, 1933) of the matrix $\Sigma^{[g]}$ is used:

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

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

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

where

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

for $j=1,\ldots,k-1$. It is important to note that $\Sigma^{[g]}$ can then be obtained from $\{\Sigma_{kk},(\Gamma_{k-1},\tau_{k-1}),\ldots,(\Gamma_1,\tau_1)\}$ by inverting 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 (GIW) model specified by:

$$\begin{cases} Y \mid \Sigma \sim N(0, I_n \otimes \Sigma); \\ \Sigma \sim GIW(\Psi, \delta), \end{cases} \tag{1}$$

where Ψ is a function of a set of parameters involved in gauged and ungauged sites, i.e. $\Psi = [\Omega, \Lambda^{[g]}, \Lambda^{[u]}, \tau_0^{[u]}, H^{[u]}]$; Ω is a $p \times p$ covariance matrix for responses at each site assumed constant from site-to-site, $\delta = [\delta^{[u]}, \delta^{[g]}]$, $\otimes$ represents the Kronecker product between matrices and GIW represents the Generalized Inverted Wishart distribution developed by Brown, Le, Zidek (1994b). The Kronecker product is chosen because it reduces the number of parameters greatly and more importantly allows us to estimate the parameters in ungauged sites i.e. census tracts where we do not have observed data. Here $N_p(0, \Theta)$ denotes the p dimensional Gaussian distribution with mean 0 and covariance matrix Θ.

The conjugate prior distribution for Σ in (1), can equivalently be presented in terms of new parameters ($\Sigma^{[g]}$, $\tau^{[u]}$, $\Gamma^{[u]}$). Specifically, $\Sigma \sim GIW(\Psi, \delta)$

$$\begin{cases} \Sigma^{[g]} \sim GIW(\eta, \delta^{[g]}); \\ \tau^{[u]} \mid \Gamma^{[u]} \sim N\left(\tau_0^{[u]}, H^{[u]} \otimes \Gamma^{[u]}\right); \\ \Gamma^{[u]} \sim IW(\Lambda^{[u]} \otimes \Omega, \delta^{[u]}), \end{cases} \tag{2}$$

where η is a function of parameters for gauged sites. That is $\eta = [\Omega, \Lambda_k, \Lambda_j, H_j, \tau_{0j}]$ ($j = 1, 2, \ldots k-1$), $\delta^{[g]} = [\delta_1, \delta_2, \ldots, \delta_k]$, $\Gamma^{[u]} = \Sigma^{[u|g]} = \Sigma^{[u]} - \Sigma^{[ug]}(\Sigma^{[g]})^{-1}\Sigma^{[gu]}$ is the residual covariance of $Y^{[u]}$, residuals after optimal linear prediction based on $Y^{[g]}$, $\tau^{[u]} = (\Sigma^{[g]})^{-1}\Sigma^{[gu]}$ is the slope of the optimal linear predictor of $Y^{[u]}$ based on $Y^{[g]}$ and $IW(\Phi, \delta)$ denotes the Inverted Wishart distribution with scale matrix Φ and degrees of freedom δ. The hypercovariance matrix $\Lambda \otimes \Omega$ is partitioned comformably 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:

$$\begin{cases} \Sigma_{kk} \sim IW(\Lambda_k \otimes \Omega, \delta_k); \\ \tau_j \mid \Gamma_j \sim N(\tau_{0j}, H_j \otimes \Gamma_j); \\ \Gamma_j \sim IW(\Lambda_j \otimes \Omega, \delta_j), \end{cases} \tag{3}$$

for $j = 1, \ldots, k-1$, H_j is a $g_j p \times g_j p$ matrix while the δ_j represent the degrees of freedom for the j^{th} block, Σ_{kk} is a $g_k p \times g_k p$, matrix, τ_j is a $(g_{j+1} + \ldots\ldots + g_k)p \times g_j p$ matrix, and Γ_j is a $g_j p \times g_j p$ matrix.

The hyperparameters involved in the GIW model are written as

$$\mathcal{H} = \left\{\Lambda^{[u]}, \delta^{[u]}, \tau_0^{[u]}, H^{[u]}, \Lambda_k, \delta_k, \Omega, [\tau_{0j}, H_j, \Lambda_j, \delta_j], j = 1, \ldots, k-1\right\}. \tag{4}$$

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 of the blocks can be expressed through the hyperparameter vector δ. More details on the characteristics of the GIW distribution are given in Brown et al (1994b), and Le et al (1999). In the following section we will present the main result of the paper.

2.3 The Posterior Distributions

Since the posterior distribution is an essential element of the predictive distribution, we develop and investigate the relevant posterior distributions in this section. Following Le et al (1999) the posterior density can be obtained and expressed by the following theorem.

Theorem 1. The joint posterior density, $f(\Sigma^{[g]} \mid D, \mathcal{H})$ is given by

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

with

$$\begin{aligned}
\Sigma_{kk} \mid D, \mathcal{H} &\sim IW(\hat{\Lambda}_k, \delta_k + n - m_k), \\
\tau_j \mid D, \Gamma_j, \mathcal{H} &\sim N\left(\hat{\tau}_j, \hat{H}_j \otimes \Gamma_j\right), \qquad (5) \\
\Gamma_j \mid D, \mathcal{H} &\sim IW(\hat{\Lambda}_j, \hat{\delta}_j),
\end{aligned}$$

where:

$$\begin{aligned}
\hat{\Lambda}_k &= \Lambda_k \otimes \Omega + (Y^{[s_k^2]})' Y^{[s_k^2]}, \\
\hat{\Lambda}_j &= \Lambda_j \otimes \Omega + \left(Y^{[s_j^2]} - Y_2^{[g_j+1,\ldots,g_k]} \tau_{0j}\right)' \\
&\times \left[I_{n-m_j} + \left(Y_2^{[g_j+1,\ldots,g_k]}\right)' H_j Y_2^{[g_j+1,\ldots,g_k]}\right]^{-1} \left(Y^{[s_j^2]} - Y_2^{[g_j+1,\ldots,g_k]} \tau_{0j}\right); \\
\hat{H}_j &= \left[H_j^{-1} + \left(Y_2^{[g_j+1,\ldots,g_k]}\right)' Y_2^{[g_j+1,\ldots,g_k]}\right]^{-1}; \\
\hat{\tau}_j &= \hat{H}_j \left[H_j^{-1} \tau_{0j} + \left(Y_2^{[g_j+1,\ldots,g_k]}\right)' \left(Y_2^{[g_j+1,\ldots,g_k]}\right)\right]; \\
\hat{\delta}_j &= \delta_j + n - m_j; \qquad (6)
\end{aligned}$$

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

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 hyperparameters estimations are given in Section 3.

2.4 The Predictive Distributions for Gauged sites

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

$$\begin{pmatrix} \mu_{(1)}^{[j]} \\ \mu_{(2)}^{[j]} \end{pmatrix} : \begin{pmatrix} m_j \times g_j p \\ (n - m_j) \times g_j p \end{pmatrix} = \begin{pmatrix} Y_1^{[g_j+1,\cdots,g_k]} \\ Y_2^{[g_j+1,\cdots,g_k]} \end{pmatrix} \tau_{0j};$$

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

$$= I_n + Y^{[g_j+1,\cdots,g_k]} H_j (Y^{[g_j+1,\cdots,g_k]})', \quad for\ j = 1, 2, \ldots, k-1.$$

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

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

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

$$\Psi_{(1|2)}^{[j]} = \frac{\Lambda_j \otimes \Omega}{\delta_j - g_j p + 1} (Y^{[g_j^t]} - \mu_{(2)}^{[j]})' (A_{22}^{[j]})^{-1} (Y^{[g_j^t]} - \mu_{(2)}^{[j]}); \tag{7}$$

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

Then following Le et al (1999), the predictive distribution of the unobserved responses conditioned on the observed data can be expressed in the following theorem.

Theorem 2. *The predictive distribution of the unobserved responses Y_{unob} conditional on the observed data D and the hyperparameter set $\mathcal{H}$ for gauged sites is given by*

$$(Y_{unob} \mid D, \mathcal{H}) \sim \prod_{j=1}^{k-1} \left(Y^{[g_j^t]} \mid Y^{[g_{j+1}^t,\cdots,g_k^t]}, D, \mathcal{H} \right) \times \left(Y^{[g_k^t]} \mid D, \mathcal{H} \right), \tag{8}$$

where the two components of the conditional distributions are specified as follows.

(i)

$$\left(Y^{[g_k^t]} \mid D, \mathcal{H} \right) \sim t_{m_k \times g_k p} \left(0, \Phi_{(1|2)}^{[k]} \otimes \Psi_{(1|2)}^{[k]}, \delta_{(1|2)}^{[k]} \right); \tag{9}$$

(ii)

$$\left(Y^{[g_j^t]} \mid Y^{[g_{j+1}^t,\cdots,g_k^t]}, D, \mathcal{H} \right) \sim t_{m_j \times g_j p} \left(\mu_{(1|2)}^{[j]}, \Phi_{(1|2)}^{[j]} \otimes \Psi_{(1|2)}^{[j]}, \delta_{(1|2)}^{[j]} \right). \tag{10}$$

We refer to the factors in (9) and (10) as *backcasting* since they give the joint predictive distribution of the response variables at the gauged sites during their ungauged time period.

Corollary 1: The means of the predictive distributions in Theorem 2 (and used in 7) are given below:

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

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*

$$E\left(\mu_{(1)}^{[j]} \mid D, \mathcal{H} \right) = \mu_{1,j}^{[g_j+1,\cdots,g_k]} \tau_{0j},$$

$$E\left(A_{12}^{[j]} \mid D, \mathcal{H} \right) = \mu_{1,j}^{[g_j+1,\cdots,g_k]} H_j \left(\mu_{2,j}^{[g_j+1,\cdots,g_k]} \right)'$$

where

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

For known hyperparameters, the unobserved response matrix follows a Matrix T distribution with appropriate covariance matrices and degrees of freedom. However, for unknown hyperparameters, the predictive distribution can be approximated by its estimates obtained by using the method of moments. To obtain an estimate for the covariance matrix with Kronecker structure is a challenging problem, which we discuss in the next section.

3 Estimation of Hyperparameters

In this section, we discuss the estimation of the hyperparameters in $\mathcal{H}$, which involves two steps. In the first step, where data are available for direct estimation, the hyperparameters are found by using the method of moments based on the available data. 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) and has discussed in Sections 4 and 5.

3.1 Method of Moment Estimation Equations

Here we discuss the estimation of hyperparameters in $\mathcal{H}$ of Equation (4). Specifically, we derive the method of moment estimating equation (MEE) corresponding to the model developed in the previous section that involves the Kronecker structure. To derive the moment estimating equations, we consider the joint distribution of $(Y_2^{[g_1,g_2,\ldots,g_k]})$ as,

$$\begin{aligned} f\left(Y_2^{[g_1,g_2,\ldots,g_k]}\right) &= \textstyle\prod_{j=1}^{k-1} f\left(Y_2^{[g_j]} \mid Y_2^{[g_{j+1},\ldots,g_k]}\right) f(Y_2^{[g_k]}) \\ &= \textstyle\prod_{j=1}^{k-1} t_{(n-m_j)\times g_j p}\left(Y_2^{[g_{j+1},\ldots,g_k]}\tau_{0j}, \frac{1}{\delta_j - g_j p+1}\left(I_{n-m_j} + \left(Y_2^{[g_{j+1},\ldots,g_k]}\right) H_j \left(Y_2^{[g_{j+1},\ldots,g_k]}\right)'\right) \otimes \Psi_j,\right. \\ &\quad \left.\delta_j - g_j p + 1\right) \times t_{n\times g_k p}\left(0, (\delta_k - g_k p+1)^{-1} I_n \otimes \Psi_k, \delta_k - g_k p + 1\right). \end{aligned} \tag{11}$$

Unconditional on $\Sigma^{[g]}$ the second moments of the marginal distribution of $Y_t^{[g]}$ are

$$var\left[Y_{kt}\right] = \frac{\Psi_k}{\delta_k - g_k p - 1};$$

$$var\left[Y_{2t}^{[g_j]}\right] = \left(\tau_{0j}' Var\left(Y_{2t}^{[g_{j+1},\ldots,g_k]}\right)\tau_{0j} + \frac{\Psi_j\left(1 + \mathrm{tr}(H_j \times Var(Y_{2t}^{[g_{j+1},\ldots,g_k]}))\right)}{\delta_j - g_j p - 1}\right);$$

$$cov\left[Y_{2t}^{[g_{j+1},\ldots,g_k]}, Y_{2t}^{[g_j]}\right] = Var\left(Y_{2t}^{[g_{j+1},\ldots,g_k]}\right)\tau_{0j},$$

where $j = 1, 2, \ldots, k-1$ and $t = m_j + 1, \ldots, n$.

Consider the following restrictions,

$$H_{j-1} = \begin{pmatrix} \Psi_j^{-1} & -\Psi_j^{-1}\tau_{0j}' \\ -\tau_{0j}\Psi_j^{-1} & H_j + \tau_{0j}\Psi_j^{-1}\tau_{0j}' \end{pmatrix}, \text{ for } j = 2, \ldots, k-1, \tag{12}$$

and $H_{k-1} = \Psi_k^{-1}$; which follows from the Bartlett decomposition of the inverted Wishart distribution, we obtain the $Var\left(Y_{2t}^{[g_j,\cdots,g_k]}\right)$ as

$$Var\left(Y_{2t}^{[g_j,\cdots,g_k]}\right) = \begin{pmatrix} \tau_{0j}' Var\left(Y_{2t}^{[g_{j+1},\cdots,g_k]}\right)\tau_{0j} + a_j\Psi_j & \tau_{0j}' Var\left(Y_{2t}^{[g_{j+1},\cdots,g_k]}\right) \\ Var\left(Y_{2t}^{[g_{j+1},\cdots,g_k]}\right)\tau_{0j} & Var\left(Y_{2t}^{[g_{j+1},\cdots,g_k]}\right) \end{pmatrix}, \tag{13}$$

where $a_j = \frac{1}{\delta_j - g_j p - 1}\prod_{i=j+1}^{k}\frac{\delta_i - 1}{\delta_i - g_i p - 1}$, for $(j = 1, 2, \ldots, k-1)$, $a_k = \frac{1}{\delta_k - g_k p - 1}$.

Setting up the moment estimating equations (MEE) along with the restrictions, $\Psi_j = \Lambda_j \otimes \Omega$ and $\tau_{0j} = \Theta_j \otimes I_p$, the proposed estimators of the hyperparameters by the moment estimation method at gauged sites are:

$$\hat{\tau}_{0j} = \hat{\Theta}_j \otimes I_p \approx \left(\frac{Y_2^{[g_{j+1},\cdots,g_k]'} Y_2^{[g_{j+1},\cdots,g_k]}}{n - m_j}\right)^{-1}\left(\frac{Y_2^{[g_{j+1},\cdots,g_k]'} Y_2^{[g_j]}}{n - m_j}\right) \tag{14}$$

$$\begin{aligned} \hat{\Lambda}_j \otimes \hat{\Omega} &\approx \frac{1}{a_j} \times \left(\left(\frac{Y_2^{[g_j]'} Y_2^{[g_j]}}{n - m_j}\right) - \left(\hat{\tau}_{0j}' Var(Y_2^{[g_{j+1},\cdots,g_k]})\hat{\tau}_{0j}\right)\right); \\ \hat{\Lambda}_k \otimes \hat{\Omega} &\approx (\delta_k - g_k p - 1)\left(\frac{Y^{[g_k]'} Y^{[g_k]}}{n}\right). \end{aligned} \tag{15}$$

Equation (15) can be expressed as

$$\Lambda \otimes \Omega \approx \begin{pmatrix} V_{11} & V_{12} & \cdots & V_{1g} \\ V_{21} & V_{22} & \cdots & V_{2g} \\ \cdots & \cdots & \cdots & \cdots \\ V_{g1} & V_{g2} & \cdots & V_{gg} \end{pmatrix}, \tag{16}$$

where V_{ij} are the p -dimensional block matrix and

$$\Lambda = \begin{pmatrix} \Lambda_1 & 0 & 0 & \ldots 0 \\ 0 & \Lambda_2 & 0 & \ldots 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & & \ldots \Lambda_k \end{pmatrix}.$$

3.2 Estimation Equations for $\delta_1, \delta_2, \ldots \delta_k$

To estimate the degrees of freedom, we suppose that $Y \sim t_{n \times p}(\mu, \delta^{-1} A \otimes \Psi, \delta)$. Then the differentiation of the log-likelihood function with respect to δ yields the following estimating equation for δ,

$$h(\delta \mid n, p, \mu, A, \Psi, Y) = 0,$$

where

$$\begin{aligned} h(\delta \mid n, p, \mu, A, \Psi, Y) &= \sum_{i=1}^{n+p} \eta\left(\frac{\delta + n + p - i}{2}\right) - \sum_{i=1}^{n} \eta\left(\frac{\delta + n - i}{2}\right) - \sum_{i=1}^{p} \eta\left(\frac{\delta + p - i}{2}\right) \\ &- \log |I_n + A^{-1}(Y - \mu)\Psi^{-1}(Y - \mu)'|, \end{aligned} \tag{17}$$

and $\eta(\cdot)$ is the digamma function.

From the joint distribution $f(Y_2^{[g_{j+1} \cdots g_k]})$, the estimating equations of $\delta_1, \delta_2, \ldots, \delta_k$ are then respectively given by

$$
\begin{aligned}
h\left(\delta_j - g_j p + 1 \mid n - m_j, g_j p, Y_2^{[g_{j+1} \cdots g_k]}\tau_{2j}, I_{n-m_j} + Y_2^{[g_{j+1} \cdots g_k]} H_j Y_2^{\prime [g_{j+1} \cdots g_k]}, \Psi_j, Y_2^{[g_j]}\right) &= 0 \\
h(\delta_k - g_k p + 1 \mid n, g_k p, 0, I_n, \Psi_k, Y^{[g_k]}) &= 0,
\end{aligned}
\tag{18}
$$

where $j = 1, 2, \ldots, k-1$.

To estimate the parameters in $\mathcal{H}$ by using the method of moment, we will use the following Lemma 1.

Lemma 1: Let V be an $gp \times gp$ matrix, Λ, an $g \times g$ matrix and Ω, an $p \times p$ matrix. Then the function

$$f(V, \Lambda, \Omega) = \| V - \Lambda \otimes \Omega \|^2$$

will be minimized, when

$$\hat{\lambda}_{ij} = \frac{\mathrm{tr}(\hat{\Omega} V_{ij})}{\mathrm{tr}(\hat{\Omega})^2} \text{ and } \hat{\Omega} = \frac{\sum_{ij}^{g} \hat{\lambda}_{ij} V_{ij}}{\sum_{ij}^{g} \hat{\lambda}_{ij}^2},$$

where V_{ij}, $i = j = 1, 2, \ldots, g$ is the p dimensional block matrix.

Proof: Re-write the function $f(V, \Lambda, \Omega)$ as follows,

$$
\begin{aligned}
f(V, \Lambda, \Omega) = \| V - A \otimes B \|^2 &= \mathrm{tr}[(V - \Lambda \otimes \Omega)'(V - \Lambda \otimes \Omega)] \\
&= \sum_{ij}^{g} \mathrm{tr}(V_{ij} - \lambda_{ij}\Omega)^2.
\end{aligned}
\tag{19}
$$

Minimizing (19) with respect to λ_{ij} (with Ω fixed) yields,

$$\hat{\lambda}_{ij} = \frac{\mathrm{tr}(\Omega V_{ij})}{\mathrm{tr}(\Omega)^2}, \quad i, j = 1, 2, \ldots, g.$$

To minimize (19) with respect to Ω, we consider

$$\sum_{ij}^{2} \mathrm{tr}(V_{ij} - \hat{\lambda}_{ij}\Omega)^2 = \sum_{ij}^{g} \hat{\lambda}_{ij}^2 \mathrm{tr}(\Omega^2) - 2\sum_{ij}^{g} \hat{\lambda}_{ij} \mathrm{tr}(V_{ij}\Omega) + \sum_{ij}^{g} \mathrm{tr}(V_{ij}^2). \tag{20}$$

Differentiating (20) with respect to Ω (with $\hat{\lambda}_{ij}$ fixed) and then equating to zero, yields

$$\hat{\Omega} = \frac{\sum_{ij}^{g} \hat{\lambda}_{ij} V_{ij}}{\sum_{ij}^{g} \hat{\lambda}_{ij}^2}. \tag{21}$$

The estimation steps for the parameters are as follow:

Step (I). Since the estimation of Θ_j does not depends on degrees of freedom, we estimate Θ_j $(j = 1, 2, \ldots, k)$ by using Lemma 1 and equation (14).

Step (II). Given the current values of $\hat{\delta}_1, \hat{\delta}_2, \ldots, \hat{\delta}_k$, estimate the parameters Λ_j, Λ_k, and Ω, by using the method of moment equations (15) and Lemma 1.

Step (III). For a given values of Λ_j $(j = 1, 2, \ldots, k-1)$, Λ_k and Ω, and estimated values of

Θ_j ($j = 1, 2, \ldots, k$), we estimate the degrees of freedom $\delta_1, \delta_2, \ldots \delta_k$ by the estimating equation (18).
Step (IV). Repeat the steps (II) and (III) until convergence.

In the following section, we will discuss about the the predictive distribution and the related parameter estimations for ungauged sites.

4 Predictive Distributions for Ungauged Sites

This section provides the joint predictive distributions of all unobserved responses at ungauged sites. Consider the distribution of Y in (1) and the prior distribution of Σ in (2). We may state the predictive distribution of unobserved responses at ungauged sites as in the following theorem.

Theorem 3. *The predictive distribution of the unobserved responses at ungauged sites conditional on the observed data $Y^{[g]}$ and the hyperparameter set $\mathcal{H}$ is given by*

$$\begin{aligned}(Y^{[u]} \mid \hat{Y}^{[g]}, \hat{\mathcal{H}}) \;\sim\; & t_{n\times ub}\Big\{\hat{Y}^{[g]} \times \hat{\tau}_0^{(u)}, (\hat{\delta}^{[u]} - up + 1)^{-1} \\ & \times \;(I_n + \hat{Y}^{[g]}\hat{H}^{[u]}\hat{Y}^{[g]'}) \otimes (\hat{\Lambda}^{[u]} \otimes \hat{\Omega}), \hat{\delta}^{[u]} - up + 1\Big\}, \end{aligned} \tag{22}$$

where $\hat{Y}^{[g]}$ matrix contains both observed and backcasted responses and $H^{[u]}$ is obtainable from (12) in terms of $[\Psi_j, \tau_j \; j = 1, 2, \ldots, k-1.]$

4.1 Estimation of Parameters for Ungauged Sites

The hyperparameters Λ_j, Θ_j ($j = 1, 2, \ldots, k-1$), δ_j ($j = 1, 2, \ldots, k$), and Ω, have already been estimated by using the method of moment in Section 3. The remaining hyperparameters related to ungauged sites, are estimable by using the SG method. Note that SG method is non-parametric and is designed to extend the spatial covariance from the gauged to the ungauged sites. To proceed to the SG algorithm, we need the following joint distribution.

The joint distribution of $\left(Y^{[u]}, Y^{[g]}\right)$ is

$$\begin{aligned} & f(Y^{[u]}, Y^{[g]} \mid \mathcal{H}) = f(Y^{[u]} \mid Y^{[g]}, \mathcal{H}) f(Y^{[g]} \mid \mathcal{H}) \\ = \; & t_{n\times up}\left(Y^{[g]}\tau_0^{[u]}, (\delta^{[u]} - up + 1)^{-1}\left[I_n + Y^{[g]}H^{[u]}Y^{[g]'}\right] \otimes \Psi^{[u]}, \delta^{[u]} - up + 1\right) \\ \times \; & \prod_{j=1}^{k-1} t_{n\times g_j p}\left(Y^{[g_{j+1}, \ldots, g_k]}\tau_{0j}, \frac{1}{\delta_j - g_j p + 1}\left(I_n + Y^{[g_{j+1}, \ldots, g_k]}H_j Y^{[g_{j+1}, \ldots, g_k]'}\right) \otimes \Psi_j, \delta_j - g_j p + 1\right) \\ \times \; & t_{n\times g_k p}\left(0, (\delta_k - g_k p + 1)^{-1} I_n \otimes \Psi_k, \delta_k - g_k p + 1\right). \end{aligned} \tag{23}$$

The $\tau_0^{[u]}$ and $\Psi^{[u]}$ are to be solved using the empirical variance matrix V and the moments of the joint distribution.

Use the following parametric restrictions: $\Psi^{[u]} = \Lambda^{[u]} \otimes \Omega$; $\tau_0^{[u]} = \Theta_0^{[u]} \otimes I_p$, with $\Theta_0^{[u]} : g \times u$ and $H^{[u]} = H_0$, where H_0 can be obtained from (12) for $j = 1, 2, \ldots, k-1$; we obtain the population covariance matrix between ungauged and gauged sites as follows

$$Var\left(Y_t^{[u]}, Y_t^{[g]}\right) \;=\; \begin{pmatrix} \tau_0^{[u]'} Var\left(Y_t^{[g]}\right)\tau_0^{[u]} + \frac{\left(1 + \mathrm{tr}(H^{[u]} \times Var(Y_t^{[g]})\right)}{\delta_u - up - 1}\Psi_u & \tau_0^{[u]'} Var\left(Y_t^{[g]}\right) \\ Var\left(Y_t^{[g]}\right)\tau_0^{[u]} & Var\left(Y_t^{[g]}\right) \end{pmatrix},$$

$$
\begin{aligned}
&= \begin{bmatrix} \Theta^{[u]'}\Lambda_1\Theta^{[u]} + \frac{1}{\delta_u - up - 1}\prod_{j=1}^{k}\frac{\delta_j - 1}{\delta_j - g_j p - 1}\Lambda^{[u]} & \Theta^{[u]'}\Lambda_{(1)} \\ \Lambda_{(1)}\Theta^{[u]} & \Lambda_{(1)} \end{bmatrix} \otimes \Omega \\
&= A \otimes \Omega, \qquad (24)
\end{aligned}
$$

where

$$
\Lambda_{(j)} = \begin{bmatrix} \Theta_j'\Lambda_{(j+1)}\Theta_j + \frac{1}{\delta_j - g_j p - 1}\prod_{i=j+2}^{k}\frac{\delta_i - 1}{\delta_i - g_i p - 1}\Lambda_j & \Theta_j'\Lambda_{(j+1)} \\ \Lambda_{(j+1)}\Theta_j' & \Lambda_{(j+1)} \end{bmatrix},
$$

for $(j = 1, 2, \ldots, k-1)$ and $\Lambda_{(k)} = \Lambda_k$.

Then SG method is used to obtain an estimate of Σ as follows:

$$
\hat{\Sigma} = \hat{V} = \begin{bmatrix} \hat{V}^{[uu]} & \hat{V}^{[ug]} \\ \hat{V}^{[gu]} & \hat{V}^{[gg]} \end{bmatrix}. \qquad (25)
$$

Equating matrix A in equations (24) and (25), we obtain the estimated hyperparameters for ungauged sites:

$$
\begin{aligned}
\hat{\Lambda}_{(1)} &= \hat{V}^{[gg]} \\
\hat{\Theta}^{[u]} &= \hat{\Lambda}_{(1)}^{-1}\hat{V}^{[gu]} = \hat{V^{[gg]}}^{-1}\hat{V}^{[gu]} \\
\hat{\tau}_0^{[u]} &= \hat{\Theta}_0^{[u]} \otimes I_p = \hat{V^{[gg]}}^{-1}\hat{V}^{[gu]} \otimes I_p \\
\hat{\Lambda}^{[u]} &= \frac{1}{a_u} \times [\hat{V}^{uu} - \hat{V}^{[ug]}\hat{V^{[gg]}}^{-1}\hat{V}^{[gu]}], \qquad (26)
\end{aligned}
$$

where $a_u = \frac{1}{\delta_u - up - 1}\prod_{j=1}^{k}\frac{\delta_j - 1}{\delta_j - g_j p - 1}$.

4.2 Selection of $\delta^{(u)}$

From the joint distribution of $f(Y^{[u]}, Y^{[g]})$ the estimation equations of $\delta^{[u]}$ are then given by

$$
h((\delta^{[u]} - \mathrm{up} + 1) \mid n, \mathrm{up}, Y^{[g]}\tau_0^{[u]}, I_p + Y^{[g]}H^{[u]}Y^{[g]'}, \Lambda^{[u]} \otimes \Omega, Y^{[u]}) = 0. \qquad (27)
$$

Note that $Y^{[u]} = Y^{[g]}\tau_0^{[u]}$, so the log-determinant in h function is equal to zero, implying that the equation (27) has no solution (Le et al 1998). Since the degrees of freedom $\delta^{(u)}$, have to selected before interpolating the unobserved data. Given our lack of a spatial model on the basis of which we could interpolate the degrees of freedom over space, we select $\delta^{(u)}$ as

$$
\hat{\delta}^{(u)} = \min(\hat{\delta}_1, \ldots \hat{\delta}_k) \quad \text{or} \quad \frac{\hat{\delta}_1 + \ldots + \hat{\delta}_k}{k},
$$

subject to condition that $\delta^{(u)} \geq \mathrm{up}$, $\hat{\delta}_1, \hat{\delta}_2, \ldots, \hat{\delta}_k$ have been estimated in section 3. For our application $(k = 2)$, we consider $\delta^{[u]} = \min(\hat{\delta}_1, \hat{\delta}_2)$, the seemingly convincing choice.

Another way of selecting $\delta^{[u]}$ can be described as follows: estimate δ from different sites, for example $\delta_1, \delta_2, \ldots, \delta_k$ from k observed sites and then interpolates these estimates by using a smoothing function over all the spatial field. To be feasible however more gauged sites are needed.

5 Application

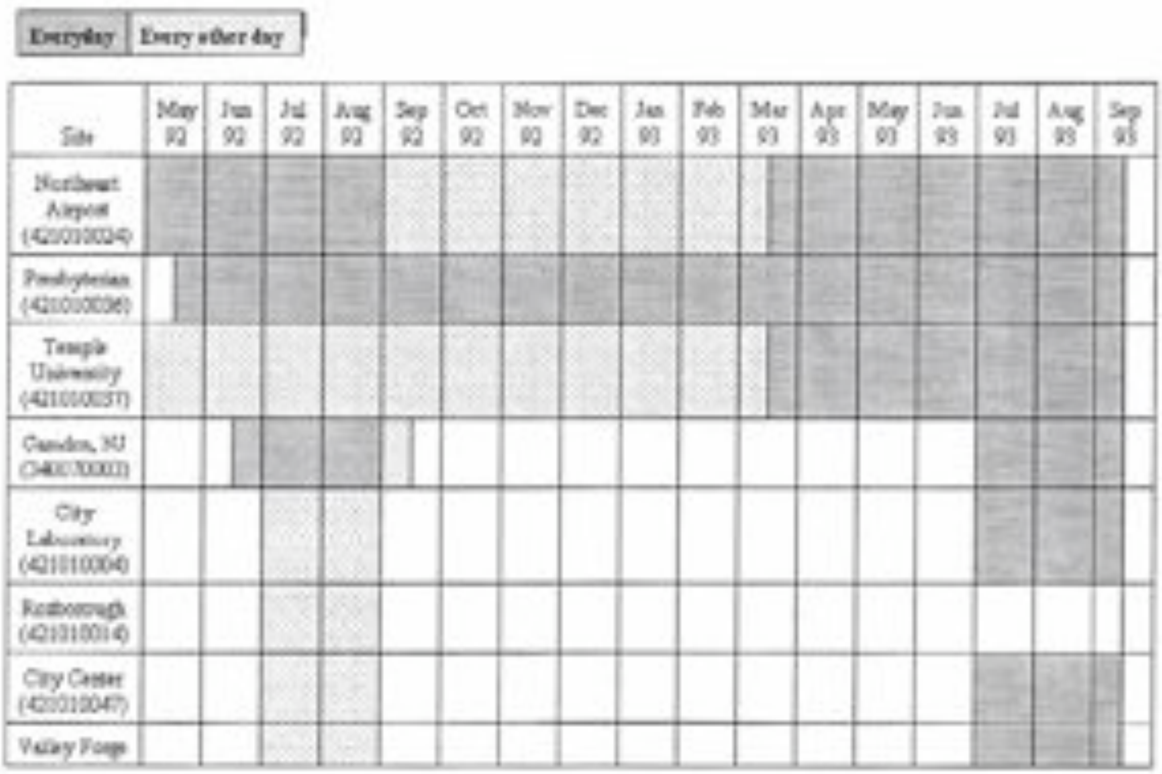
In this section we illustrate our theory using data from Philadelphia, where multivariate pollutants ($PM_{2.5}$ and PM_{10}) were measured at 8 sites, but not all sites measured all pollutants at the same time.

Le et al (1999) present a Bayesian interpolation method to deal with univariate staircase data, while a single pollutant is of interest. The method developed in this paper deals with a multivariate setting and enables to borrow strength by exploiting correlation between pollutants.

5.1 Data Patterns in Staircase

The daily average concentration levels for $PM_{2.5}$ and PM_{10} were measured (in $\mu g/m^3$) at eight (8) monitoring stations: Temple University, Presbyterian, City Center, Camden, City Laboratory, Roxborough, NE Airport and Valley Forge in Philadelphia from May 92 to the middle of September 1993. However, not all sites measured pollutants at the same time. Table 1 provides the locations of these sites, their corresponding code numbers and operating times.

Table 1. *The pattern of $PM_{2.5}$ and PM_{10} data in Philadelphia*



Observe in Table 1 that stations at Temple University, Presbyterian, and Airport have recorded concentration levels from May 1992 to the middle of September 1993. Among these three sites, Presbyterian recorded concentration levels everyday from the middle of May 1992 to September 1993. Temple University recorded concentration levels from May 1992 to the middle of March 1993 every other day and everyday from the middle of March to September in 1993. NE Airport recorded daily concentration levels from May to August in 1992 and from the middle of March to middle of September in 1993, and every other day in between these times. The other four sites (City Center, Camden, City Laboratory, and Valley Forge) reported levels every other day from July to August in 1992 and daily from July to the middle of September in 1993. Among four sites, the only exception is observed in Camden station which recorded daily concentration level from middle of June to August in 1992 and from July to middle of September in 1993 and every other day for two weeks in September 1992. To avoid the complexity with little loss of information, concentrations measured in June and September in 1992 at Camden station were omitted. Due to lack of measurements, we set the Roxborough site aside, its data to be used later for assessing interpolated values. We used the other seven sites to build our interpolator. The geographical locations of monitoring sites in Philadelphia is given in Figure 1.

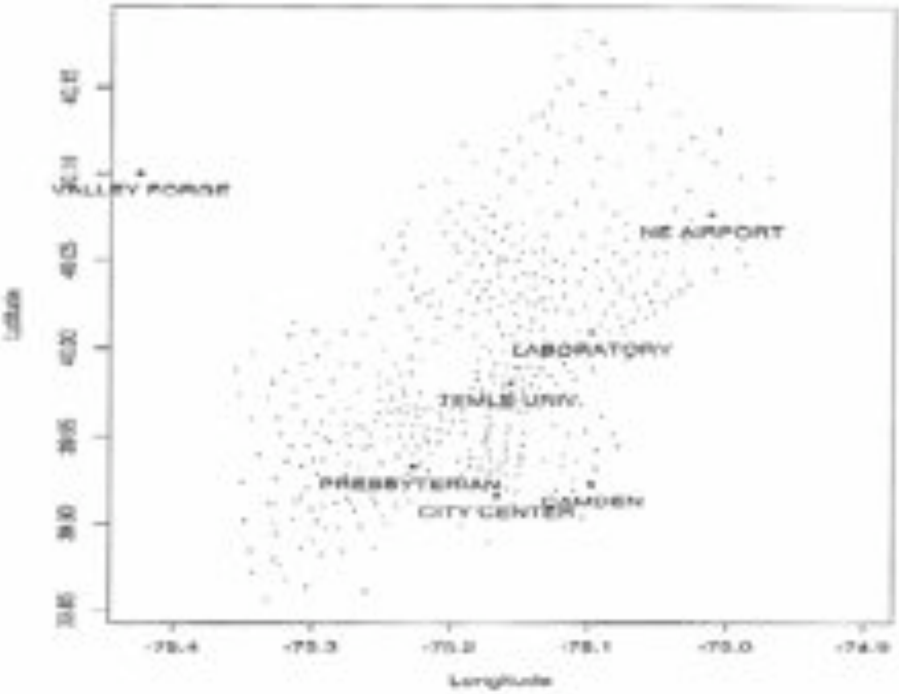


Figure 1. *Geographical locations of monitoring sites in Philadelphia. Sites are indicated by "•" and 482 census tracts plus the Roxborough site are indicated by "·".*

The unmeasured pollutants for the every other day have been filled-in using the EM algorithm. With these filled-in values we have three "complete" sites to build the second block and four other sites having the same number of missing data values to build the first block at the gauged sites.

It is noted that our theory works only for an equal number of observations in each block while the number of observations can be different between blocks. After detrending and autoregression of the logarithm of $PM_{2.5}$ and PM_{10} using the time-space model, the residuals become time independent. We therefore can move the data from July and August in 1992 to May and June in 1993 in both blocks. By arranging the data in this way we achieve a pattern that resembles a staircase. The sites with shortest records form the lowest step while those sites with longest records form the highest step. We refer to the resulting blocked data set as the "residual staircase" data. The total number of observations in block 1 is $n_1 = 137$ and in block 2 is $n_2 = 503$.

The second block contains more observations than the first one, which creates the data stair case structure and therefore, we can implement our theory on this staircase data set.

5.2 GIW Models and Backcasting

In this subsection, we apply our theory to the staircase residual data. We have assumed a Gaussian-Generalized Inverted Wishart (GIW) model. To assess that assumption we checked the staircase residuals of both PM_{10} and $PM_{2.5}$ individually by using Q-Q normal plots. The Q-Q plots produced by using residuals from each monitoring sites are shown in Figure 2. These plots make the Gaussian assumption quite tenable.

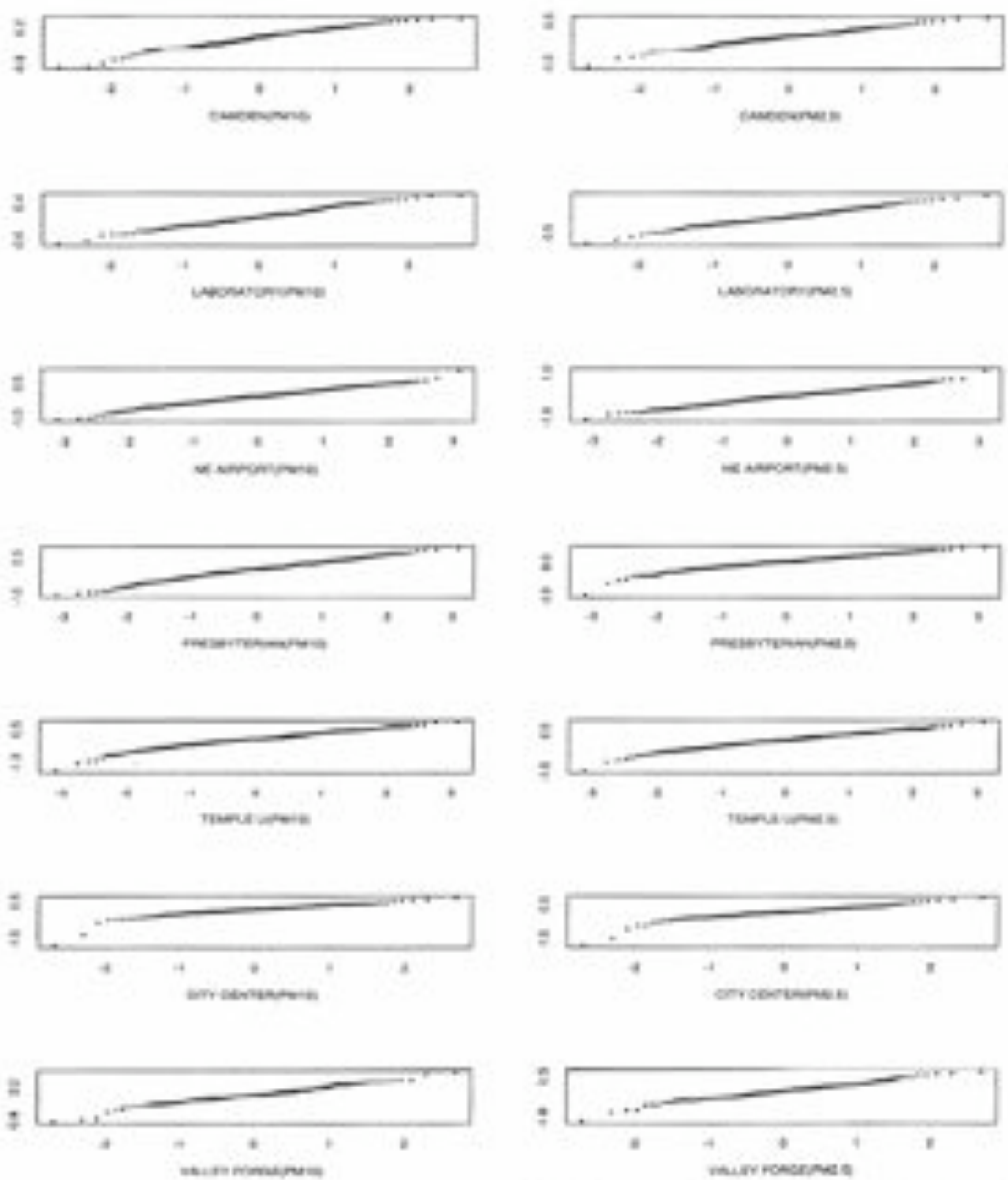


Figure 2. *Q-Q plots for "staircase" residuals of PM_{10} and $PM_{2.5}$ of all monitoring sites.*

The predictive distributions and related procedures for estimating hyperparameters were presented in Section 2. For the gauged sites we have $k = 2$, $g_1 = 4$, $g_2 = 3$, $p = 2$, $m_1 = 366$ and $n = 503$. The estimated hyper-covariance matrix corresponding to the spatial cross-correlation of the residuals at the gauged sites are as follows:

$$\hat{\Lambda}_1 = \begin{pmatrix} 5.579 & 1.428 & 2.759 & 1.767 \\ 1.428 & 6.158 & 2.419 & 2.042 \\ 2.759 & 2.419 & 15.668 & 0.459 \\ 1.767 & 2.042 & 0.459 & 6.321 \end{pmatrix}, \quad \hat{\Lambda}_2 = \begin{pmatrix} 27.832 & 21.635 & 20.145 \\ 21.635 & 30.401 & 24.323 \\ 20.145 & 24.323 & 26.978 \end{pmatrix}$$

$$\hat{\Omega} = \begin{pmatrix} 1.000 & 0.894 \\ 0.894 & 1.209 \end{pmatrix}, \quad \text{and} \quad \hat{\Theta}_0 = \begin{pmatrix} 0.503 & 0.269 & -0.072 & 0.069 \\ -0.108 & 0.015 & 0.294 & 0.225 \\ 0.085 & 0.275 & 0.339 & 0.224 \end{pmatrix}.$$

The estimated degrees of freedom at gauged sites are $\hat{\delta}_1 = 207$ and $\hat{\delta}_2 = 239$.

These estimates can be interpreted as follows. The diagonal elements of $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ are the variances of the monitoring sites in the first and second blocks respectively after jointly accounting for both $PM_{2.5}$ and PM_{10}. The variances of the sites, Camden, Laboratory, City Center and the Valley Forge are smaller in scale than the sites, NE Airport, Presbyterian and Temple University. It is also observed that City Center is generally less correlated with other sites. More of this will be discussed in a later section. The covariance matrix $\hat{\Omega}$ gives the correlations between pollutants. All the estimated parameters, $\hat{\delta}_1$, $\hat{\delta}_2$, $\hat{\Lambda}_1$, $\hat{\Lambda}_2$ and $\hat{\Omega}$ will yield the $\hat{V}_{gg}$ matrix for gauged sites.

The means of the predictive distribution at gauged sites impute the systematically missing values and this process is called the "backcasting". After backcasting, we shifted the original data from May to June in 1993 to July to August in 1992. Figure 3 shows the backcasted residuals for both $PM_{2.5}$ and PM_{10} (light-dotted lines) along with the observed concentration levels (dark-dotted lines) for both pollutants. We see that both pollutants have similar backcasted data due to the apparent "flatness" of PM fields in Philadelphia.

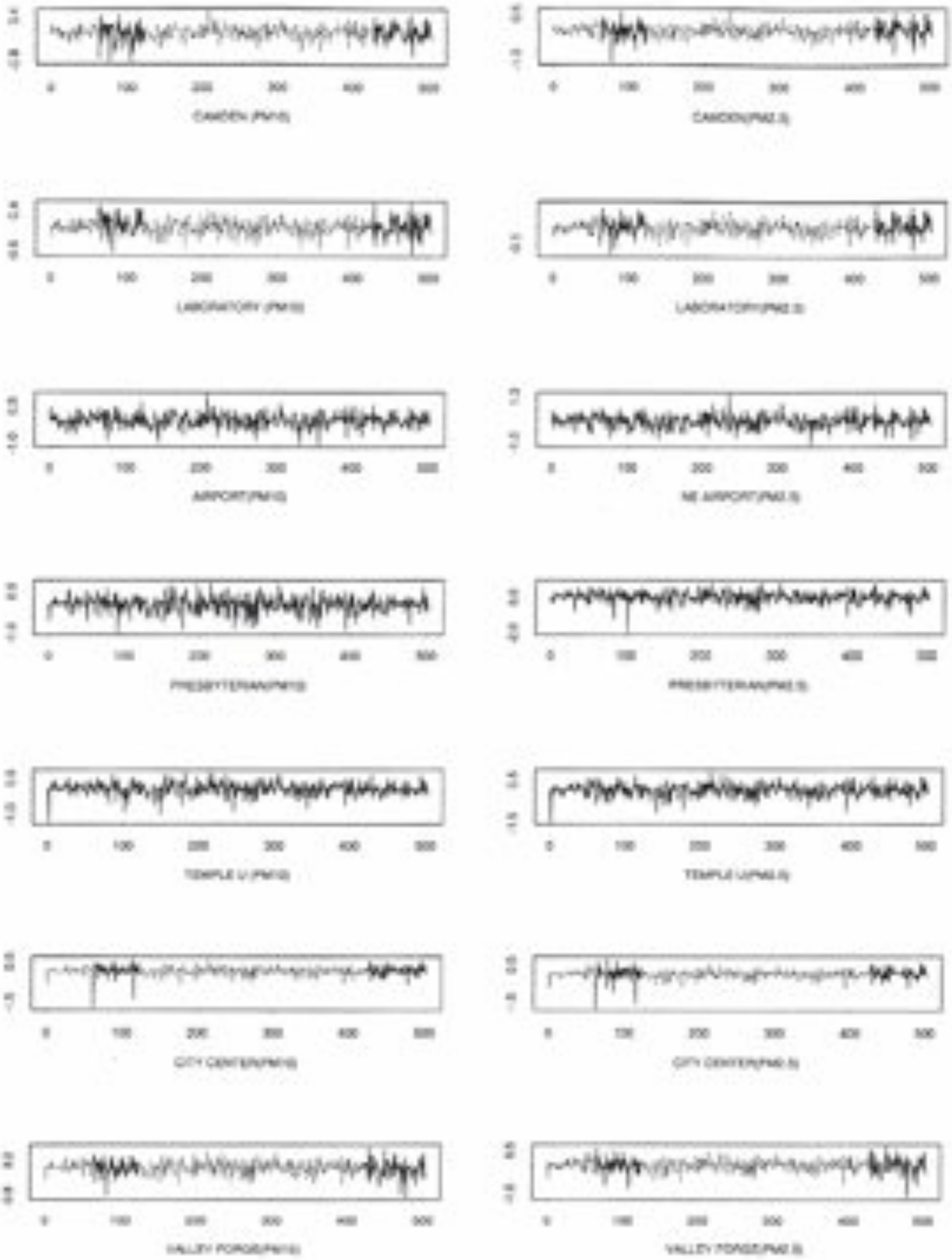


Figure 3. *Backcasted data for $PM_{2.5}$ and PM_{10} from all gauged sites.*

The estimated correlations between sites by the Kronecker product are given in Table 2. Under the assumed Kronecker correlation structure, we found that the correlation ranges between 0.401 to 0.849. The highest estimated correlation is between Presbyterian and Temple University, while the lowest is between City Center and NE Airport. We also observe that City Center has less correlation with others sites than the sites themselves.

Table 2. *Cross-correlation between monitoring sites in Philadelphia*

	Camden	Laboratory	Airport	Presbyterian	Temple	City Center	Valley
Camden	1.000	0.580	0.695	0.481	0.517	0.444	0.536
Laboratory	0.580	1.000	0.655	0.609	0.658	0.503	0.632
Airport	0.695	0.655	1.000	0.744	0.735	0.401	0.563
Presbyterian	0.481	0.609	0.744	1.000	0.849	0.536	0.657
Temple Univ.	0.517	0.658	0.735	0.849	1.000	0.541	0.653
City Center	0.444	0.503	0.401	0.536	0.541	1.000	0.406
Valley	0.536	0.632	0.563	0.657	0.653	0.406	1.000

The correlation coefficient between pollutants is 0.81. This indicates that a major contribution to PM_{10} concentration levels comes from $PM_{2.5}$.

5.3 GIW Models and Interpolation

In this subsection we discuss the predictive distribution for unobserved daily average $PM_{2.5}$ concentration levels for given observed and backcasted data in seven sites in Philadelphia based on the theory presented in Section 4. The estimated covariance matrix of the GIW model is given by

$$\hat{V}_{gg} = \begin{bmatrix} 0.059 & 0.036 & 0.038 & 0.033 & 0.059 & 0.043 & 0.043 \\ 0.036 & 0.066 & 0.045 & 0.041 & 0.059 & 0.057 & 0.058 \\ 0.038 & 0.045 & 0.123 & 0.036 & 0.049 & 0.069 & 0.065 \\ 0.033 & 0.041 & 0.036 & 0.064 & 0.050 & 0.061 & 0.057 \\ 0.059 & 0.059 & 0.049 & 0.050 & 0.123 & 0.095 & 0.089 \\ 0.043 & 0.057 & 0.069 & 0.061 & 0.095 & 0.134 & 0.107 \\ 0.043 & 0.058 & 0.065 & 0.057 & 0.089 & 0.107 & 0.119 \end{bmatrix}.$$

For the current section, the analysis is conducted by using the SG methodology for estimating the covariance structure over the region under study. The SG method is used in conjunction with latitude and longitude expand the covariance matrix $\hat{V}_{gg}$ (between observed sites) to encompass both gauged and ungauged sites. The expansion of $\hat{V}_{gg}$ to $\hat{V}_{gu}$ (the covariance between gauged and ungauged sites), $\hat{V}_{uu}$ (the covariance of ungauged sites) and $\hat{V}_{ug}$ (the covariance between ungauged and gauged sites) will be used to estimate the parameters at ungauged sites and hence to obtain the mean of the predictive distribution. The SG method develops a smooth mapping between "geographic space" (G-space) and "dispersion space" (D-space). The mapping transforms (stretches, compresses, and rotates) the latitude and longitude system of G-space into a coordinates system, D-space, in such a way that the correlation between sites depends only on the distance by which the sites are separated. This condition called isotropy is needed to allow reliable interpolation. The fitted variogram and D-space coordinates obtained by means of a spline smoothing technique are presented in Figure 4 and 5. Details of the SG methodology are available in Sampson and Guttorp (1992) and recently in Li et al (2000) among other sources.

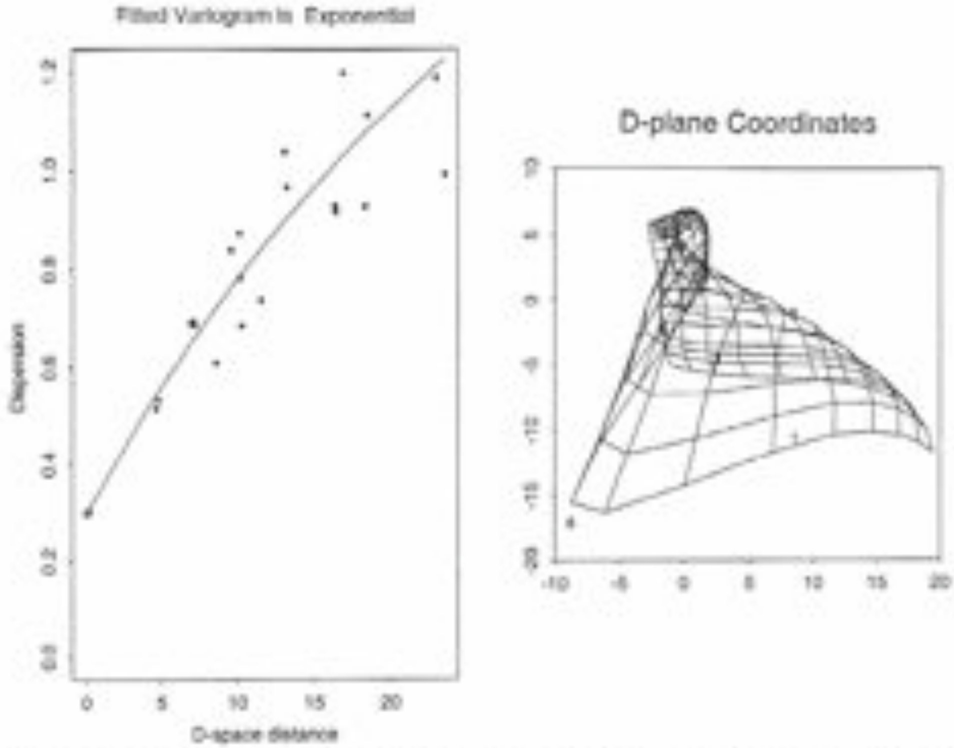


Figure 4. *Variogram Fits and D-plane: Smoothing Parameter* (λ) = 0.

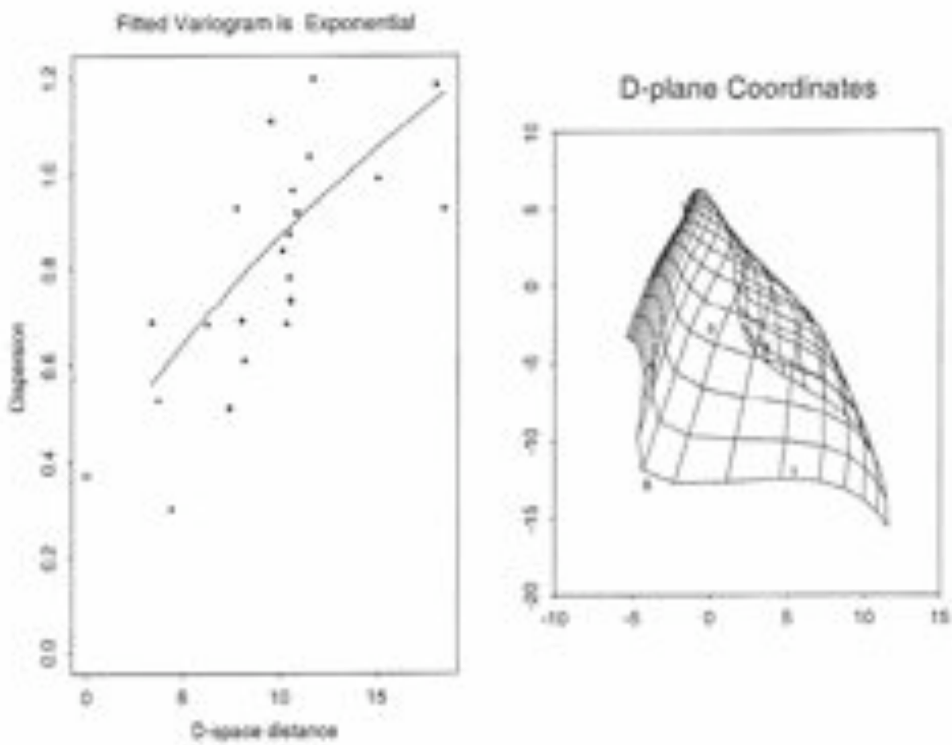


Figure 5. *Variogram Fits and D-plane: Smoothing Parameter* (λ) = 2.

In implementing 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 transform the coordinates of two locations in G-space to D-space and then compute the isotropic correlation between the transformed locations in D-space by the fitted variogram function. The left hand panel in Figures 4 and 5 show an exponential variogram fitted to the 21 D-space squared distance between the sites under investigation. For a better understanding of these plots, the movement of sites from G-space to D-space is provided in Figure-6. Figure 4 shows the surface must essentially be folded over on itself to achieve the desired states of isotropy. Figure 5 is a smoother version of Figure 4 (with smoothing parameter $\lambda = 2$). From these two figures, we may conclude that the spatial distribution of the particulates is not isotropic. To attain an isotropic state the site, NE Airport, has to be shifted towards the Center of the city, Camden has to be moved West while Valley Forge has to be moved East. Figure 6 shows another view of the movement between G-space and D-space. Since the monitoring sites, Presbyterian and Temple University are highly correlated (0.85), in D-space, one is almost superimposed on the other.

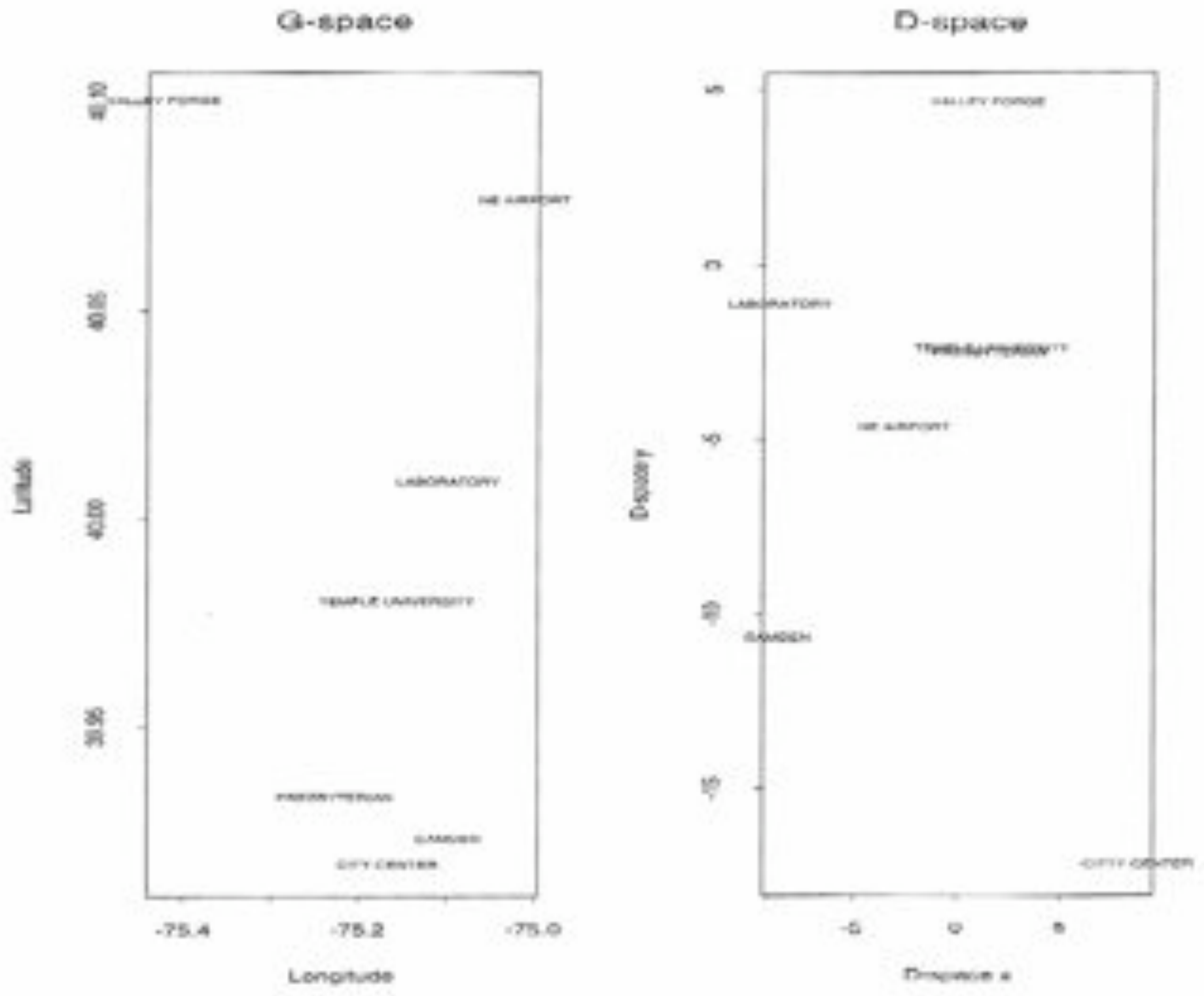


Figure 6. *Movements of sites from G-space to D-space.*

Following the discussion in Li et al (2000), the interpolated covariance matrix is obtained via the SG method. We are then able to get the interpolated average $PM_{2.5}$ concentration levels for 483 census tracts from 1992 to 1993. This yields a 503×483 matrix of interpolants.

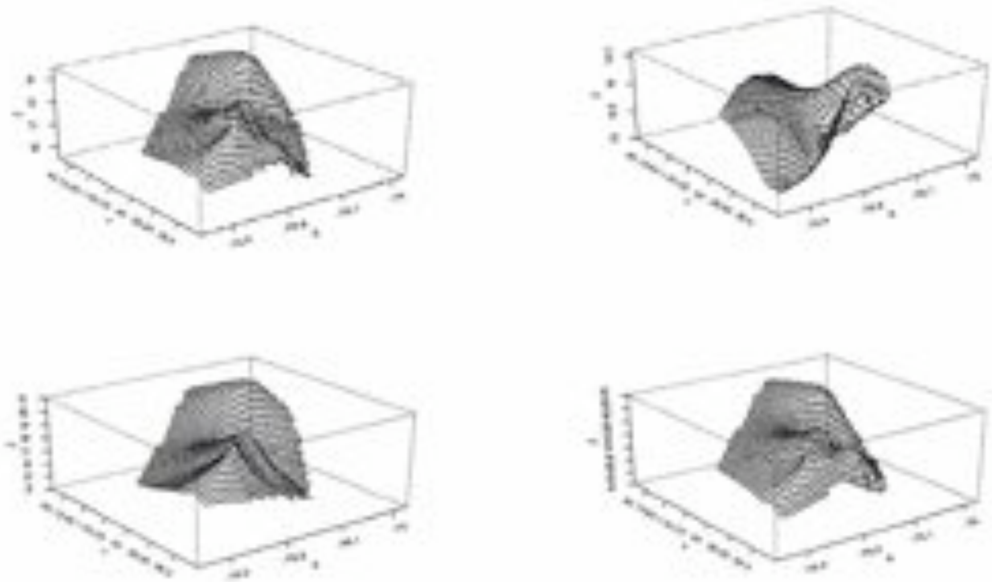


Figure 7. *Interpolated $PM_{2.5}$: January 6 (Wednesday) to 9 (Saturday) (Days 251-254).*

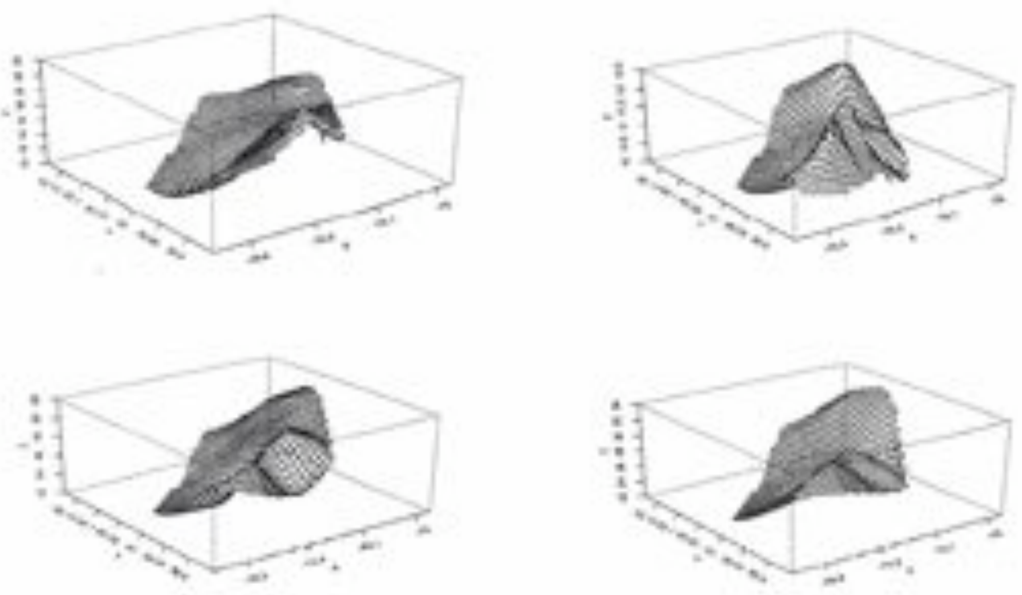


Figure 8. *Interpolated $PM_{2.5}$: September 3 (Friday) to 6 (Monday)(Days 491-494).*

With the choice of smoothing parameter $\lambda = 0$, we can apply the spatial prediction methodology discussed earlier. Figure 7 displays the interpolated daily $PM_{2.5}$ field in typical winter days

[January 6 (Wednesday) to 9 (Saturday)] in Philadelphia. Observe that the average concentration level on weekdays is higher than the weekend. Figure 8 exhibits the interpolated daily $PM_{2.5}$ field in typical summer days [September 3 (Friday) to 6 (Monday)]. Similar conclusions can be drawn for that season. However, the average daily concentration level on summer is higher than winter. In particular these figures show spatial variation in the level of $PM_{2.5}$ field over space, time (day-to-day) and over season (summer and winter). Finally, it may be concluded that the interpolated $PM_{2.5}$ surface is not flat on any day.

The standard deviation plot for the interpolated log $PM_{2.5}$ of day 493 is shown in Figure 9.

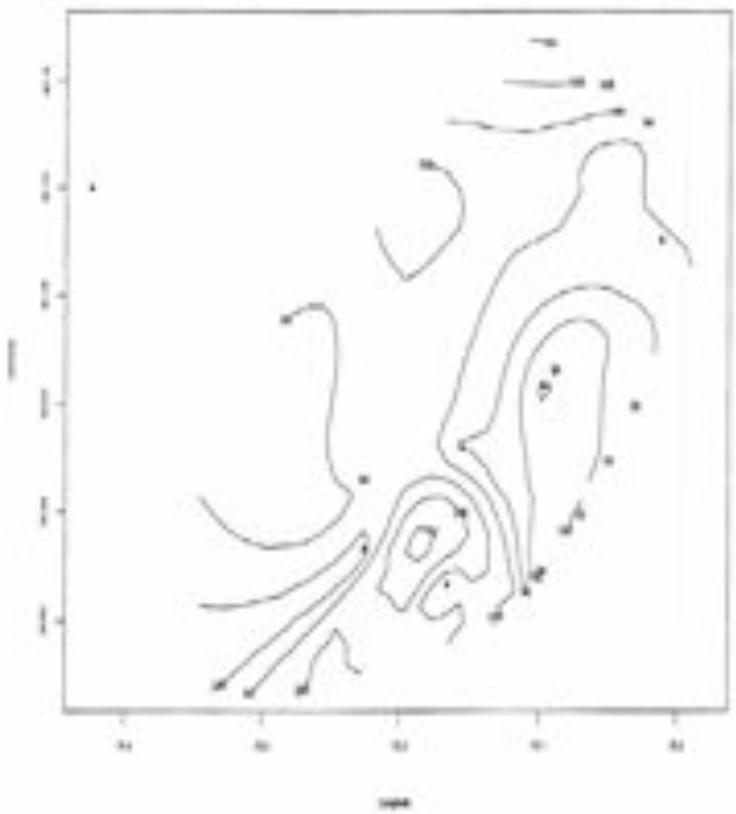


Figure 9. *Contour Plot of The SDs of The Interpolation: Day 493.*

The contour plot shows that an ungauged site close to an existing (gauged) site has a small SD and the site further away from the gauged sites has bigger SD.

5.4 Assessment of Interpolation

Since the interpolated spatial fields can serve a variety of important purposes, high predictive accuracy is desirable. In this section we will assess about the accuracy of interpolated values. It is noted that the predictive distribution of unobserved responses at ungauged sites has a matric t distribution. The marginal distribution of a matric t is again a matric t distribution (see Press 1982). When the number of rows or columns is one, the matric t become a multivariate

Student t distribution. From a multivariate Student t distribution, one can derive the univariate predictive distribution for any pollutant at any given site. Then using the t distribution, one can construct a 95% confidence band for unobserved responses. We use the estimated degrees of freedom at ungauged site as 207, which is the minimum of the estimated degrees of freedoms at gauged sites ($\hat{\delta}_1 = 207$ and $\hat{\delta}_2 = 239$). Figure 10 depicts the observed value (with big "·") at Roxborough station and the 95% confidence band (dotted line) based on the interpolated values from other seven sites. We found that the estimated 95% confidence band covered 85% (5 out of 32 data points are outside of the confidence band) while the estimated 99% confidence band covered 94% (2 out of 32 data points are outside of the confidence band) of the observed values at the Roxborough monitoring site

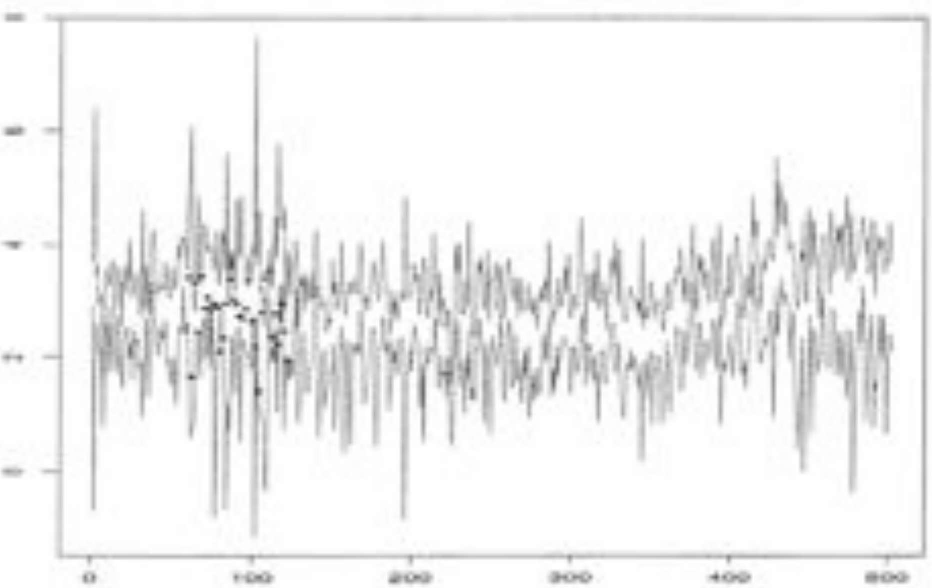


Figure 10. *Confidence band and the observed $PM_{2.5}$ at Roxborough monitor site.*

6 Concluding Remarks

This paper has developed a Bayesian approach for multivariate spatial and temporal interpolation problem. This approach is an extension of the Bayesian methodology for spatial interpolation developed by Le, Sun and Zidek (1999) to gain an interpolation theory for multiple pollutants. We assume a Gaussian generalized inverted Wishart (GIW) model. Specifically, the responses are assumed to follow a Gaussian distribution and the corresponding covariance is assumed to follow a generalized inverted Wishart prior distribution. The resulting predictive distribution follows a matrix T distribution with appropriate covariance parameters and degrees of freedom. As an alternative to the EM algorithm, we have presented in the paper a simpler method of moment estimation tool to estimate the unknown hyperparameters for multiple response models. The MOM is a straight forward method and gives exact solution. We also developed a predictive distribution for the unobserved responses at ungauged sites. The results obtained in this paper will allow us to analyse the data from different sites as well as multiple pollutants, where the observed data monitoring stations follow a staircase structure. The approach works well in the interpolation of $PM_{2.5}$ concentration levels in Philadelphia.

ACKNOWLEDGEMENTS

This research is partially supported by grants from Environmental Protection Agency, USA and the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Bates, D. V., Baker-Anderson, M and Sizto, R (1990). Asthma attack periodicity: a study of hospital emergency visits in Vancouver. *Environmental Research*, 51, 51-70.

Brown, P. J., Le, N. D., and Zidek, J. V. (1994a). Multivariate spatial interpolation and exposure to air pollutants. *Can J. Statist.*, 22, 489-510.

Brown, P. J., Le, N. D., and Zidek, J. V. (1994b). Inference for a covariance matrix. In Smith , AFM and Freedman, PR eds. *Aspects of Uncertainty. A Tribute to D. V. Lindley.* Wiley.

Burnett, R. T., Dales, R. E., Raizenne, M. R., Krewski, D., Summers, P. W., Roberts, G. R., Raad-Young, M., Dann, T., Brook, J. (1994). Effects of low ambient levels of ozone and sulfates on the frequency of respiratory admissions to Ontario hospitals. *Environmental Research*, 65, 172-94.

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

Cressie, C. F. (1991). Statistics for Spatial Data. Wiley.

Cui, H., Stein, A., and Myers, D. E. (1995). Extension of spatial information, Bayesian Kriging and updating of prior variogram parameters. *Environmetrics*, 4, 373-384.

De Oliveira, V., Kedem, B., and Short, D. A. (1997). Bayesian prediction of transformed Gaussian random fields. *J. Amer. Statist. Assoc.*, 92, 1422-1433.

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

Dockery, D. W., Pope CA III, Xu, X. (1993). An association between air pollution and mortality in Six US Cities. *N. Engl. J. Med.*, 329, 1753-9.

Dockery, D. W., Schwartz, J. and Spengler, J. D. (1992). Air pollution and daily mortality: associations with particulates and acid aerosols. *Environmental Research*, 59, 362-373.

Duddek, C., Le, N. D. , Zidek, J. V. , and Burnett, R. T. (1995). Multivariate Imputation in Cross Sectional analysis of health effects Associated with Air Pollution (with discussion). *Environ. Ecolog. Statist.*, 2, 191-212.

Gaudard, M., Karson, M., Linder, E. and Sinha, D. (1999). Bayesian spatial prediction, *Environ. Ecolog. Statist.*, 6, 147-171.

Handcock, M. S. and Stein, M. L. (1993). A Bayesian analysis of Kriging. *Technometrics*, 35, 403-410.

Kibria, B. M. G. (2000). Multivariate Bayesian spatial interpolation using Gaussian-Generalized inverted Wishart Model, Technical Report # 191, Dept. of Statistics, The University of British Columbia, Vancouver, Canada.

Kitanidis, P. K. (1986). Parameter uncertainty in estimation of spatial functions: Bayesian Analysis, *Water Resources Research*, 22, 497-507.

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

Le, N. D., Sun, W. and and Zidek, J. V. (1997). Bayesian Multivariate Spatial Interpolation with Data missing by design. *J. Royal Statist. Soc*, Series-B, 59, 501-510.

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

Le, N. D., Sun, L. and and Zidek, J. V. (1999). Bayesian spatial interpolation and backcasting using Gaussian-generalized interted wishart Model. TR # 185, Dept. of Statistics, The University of British Columbia, Vancouver, Canada.

Levy, J. L., Houseman E. A., Ryan, L., Richarson, D. and Spengler, J. D. (2000). Particle Concentrations in Urban Microenviornments. *Unpublished Manuscript*.

Li Sun, Zidek, J. V., Nhu D Le, and Halûk Ozkäynäk (2000). Interpolating Vancouver Daily Ambient PM_{10} field. *Environmetric* (To be appeared).

Liu, C. (1999). Efficient ML estimation of the multivariate normal distribution from incomplete data. *J. Mult. Anal.*, 69, 206-217.

Meiring, W, Monestiez, P, Sampson, P. D., and Guttorp, P. (1997). Developments in the modeling of nonstationary spatial covariance structures from space-time monitoring in data. Geostatistics, Wollongong '96, vol. 1, pp 162-173. *E Y Baasfi and N Schofield Eds.*, Kluwer Academic Publishers.

Ostro, B. D., Lipsett, M. J., Wiener, M. B. and Selner, J. C. (1991). Asthmatic responses to airborne acid aerosols. *American Journal of Public Health*, 81, 694-702.

Pope, C. A., Thun, M. J., Namboodiri, M. M., Dockery, D. W., Evans, J. S., Speizer, F. E. and Heath, C. W. (1995). Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *American Journal of Respiratory and Critical Care Medicine*, 151, 669-674.

Press, S. J. (1982). Applied Multivariate Analysis. HRW Inc, New York.

Roemer, W, Hoek, G and Brunekreef, B (1993). Effect of ambient winter air pollution on respiratory health of children with chronic respiratory symptoms. American Review of Respiratory Diseases, 147, 118-124.

Rubin, D. B. and Shaffer J. L. (1990). Efficiently Creating Multiple Imputations for Incomplete Multivariate Normal Data. *In Proceedings of the Statistical Computing Section, American Statistical Association*, 83-88.

Sampson, P. and Guttorp, P. (1992). Non-parametric estimates of non stationary spatial covariance structure. *J. Amer. Stat. Assoc.*, 87, 108-119.

Samian, D, Sampson, P. D. and Guttorp, P. (2000). Bayesian Estimation of Semi-parametric non-stationary spatil covariance structures. NRCSE-TRS No. 039., NRCSE, U of Washington.

Schmitt. A. M. amd O'Hagan, A. (2000). Bayesian Inference for nonstationary spatial covariance structure via spatial deformations. Unpublished manuscript.

Schwartz, J, Slater, D, Larson, T. V., Pierson, W. E., Koenig, J. Q. (1993). Particulate air pollution and hospital emergency visits for asthma in Seattle. *American Review of Respiratory Diseases*, 147, 826-831.

Smith, R. L. (1996). Estimating nonstationary spatial correlations. *Unpublished Manuscript.*

Smith R. L, Kim, Y Fuentes, M and Spitzner, D. (2000). Threshold dependence of mortality effects for fine and coarse particles in Phoenix, Arizona. *Journal of the Air and Waste Management Association, PM2000 issu.* Submitted.

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

Appendix. Estimation by EM Algorithm

In this section, we discuss how the elements of the hyperparameter set $\mathcal{H}$ of Equation (4) are to be estimated bt EM algorithm. Note that we need estimate only those parameters involved in the gauged sites. 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) and discussed in Section 4. The EM approach has previously been discussed by various researchers. However, Dempster, et al (1977), Chen (1979) and very recently Le et al (1999) as well as Liu (1999) are notable among others.

The EM iterative algorithm requires at iteration $p+1$, in the "E-step" the computation of

$$\begin{aligned}\mathcal{L}(\mathcal{H} \mid \mathcal{H}^{(p)}) &= E\left(\log[f(Y, \Sigma^{[g]} \mid \mathcal{H})] \mid D, \mathcal{H}^{(p)}\right) \\ &= E\left[\log f(Y|\Sigma^{[g]}) \mid D, \mathcal{H}^{(p)}\right] + E\left[\log f(\Sigma^{[g]} \mid \mathcal{H}) \mid D, \mathcal{H}^{(p)}\right],\end{aligned} \tag{28}$$

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

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

$$\mathcal{L}^*(\mathcal{H} \mid \mathcal{H}^{(p)}) = E\left[\log f(\Sigma^{[g]} \mid \mathcal{H}) \mid D, \mathcal{H}^{(p)}\right] \tag{29}$$

at the E-step and maximize $\mathcal{L}^*$ over $\mathcal{H}$ at the M-step, where $f(\Sigma^{[g]} \mid \mathcal{H})$ is given by

$$\begin{aligned} f(\Sigma^{[g]}|\mathcal{H}) \;\propto\; & |\Lambda_k \otimes \Omega|^{\frac{\delta_k}{2}} |\Sigma_{kk}|^{-\frac{\delta_k + g_k + 1}{2}} e^{-\frac{1}{2}\mathrm{tr}((\Lambda_k \otimes \Omega)\Sigma_{kk}^{-1})} \prod_{j=1}^{k-1} \Big\{ c(\delta_j, g_j, p) |\Lambda_j \otimes \Omega|^{\frac{g_j p}{2}} \\ \times\; & |\Gamma_j|^{-\frac{(g_{j+1} + \ldots + g_k)p}{2}} e^{-\frac{1}{2}\mathrm{tr}\left\{(\Lambda_j \otimes \Omega)(\tau_j - \tau_{j0})\Gamma_j^{-1}(\tau_j - \tau_{j0})'\right\}} |\Lambda_j \otimes \Omega|^{\frac{\delta_j}{2}} |\Gamma_j|^{-\frac{\delta_j + g_j p + 1}{2}} \\ \times\; & e^{-\frac{1}{2}\mathrm{tr}\left\{(\Lambda_j \otimes \Omega)\Gamma_j^{-1}\right\}} \Big\}. \end{aligned}$$

Then the $\log f(\Sigma^{[g]}|\mathcal{H})$ is obtained as

$$\begin{aligned} \log f(\Sigma^{[g]}|\mathcal{H}) \;\propto\; & \sum_{j=1}^{k} c(\delta_j, g_j, p) + \frac{\delta_k}{2} \log|\Lambda_k \otimes \Omega| - \left(\frac{\delta_k + g_k + 1}{2}\right) \log|\Sigma_{kk}| - \frac{1}{2} tr\{(\Lambda_k \otimes \Omega)\Sigma_{kk}^{-1}\} \\ + & \sum_{j=1}^{k-1} \frac{g_j p}{2} \log|\Lambda_j \otimes \Omega| - \sum_{j=1}^{k-1} \frac{(g_{j+1} + \ldots + g_k)p}{2} \log|\Gamma_j| \\ - & \frac{1}{2} \sum_{j=1}^{k-1} \mathrm{tr}\left\{(\Lambda_j \otimes \Omega)(\tau_j - \tau_{j0})\Gamma_j^{-1}(\tau_j - \tau_{j0})'\right\} \\ + & \sum_{j=1}^{k-1} \frac{\delta_j}{2} \log|\Lambda_j \otimes \Omega| - \sum_{j=1}^{k-1} \left(\frac{\delta_j + g_j p + 1}{2}\right) \log|\Gamma_j| - \sum_{j=1}^{k-1} \frac{1}{2} \mathrm{tr}\{(\Lambda_j \otimes \Omega)\Gamma_j^{-1}\} \end{aligned}$$

Hence,

$$\begin{aligned} \mathcal{L}^*(\mathcal{H} \mid \mathcal{H}^{(p)}, D) \;\propto\; & \sum_{j=1}^{k} c(\delta_j, g_j, p) + \frac{\delta_k}{2} \log|\Lambda_k \otimes \Omega| - \left(\frac{\delta_k + g_k + 1}{2}\right) E\left(\log|\Sigma_{kk}| \mid D, \mathcal{H}\right) \\ - & \frac{1}{2} \mathrm{tr}\{(\Lambda_k \otimes \Omega) E\left(\Sigma_{kk}^{-1} \mid D, \mathcal{H}\right)\} + \sum_{j=1}^{k-1} \frac{g_j p}{2} \log|\Lambda_j \otimes \Omega| \\ - & \sum_{j=1}^{k-1} \frac{(g_{j+1} + \ldots + g_k)p}{2} E\left(\log|\Gamma_j| \mid D, \mathcal{H}\right) \\ - & \sum_{j=1}^{k-1} \frac{1}{2} \mathrm{tr}\left\{(\Lambda_j \otimes \Omega) E\left((\tau_j - \tau_{j0})\Gamma_j^{-1}(\tau_j - \tau_{j0})' \mid D, \mathcal{H}\right)\right\} \\ + & \sum_{j=1}^{k-1} \frac{\delta_j}{2} \log|\Lambda_j \otimes \Omega| - \sum_{j=1}^{k-1} \frac{\delta_j + g_j p + 1}{2} E\left(\log|\Gamma_j| \mid D, \mathcal{H}\right) \\ - & \sum_{j=1}^{k-1} \frac{1}{2} \mathrm{tr}\{(\Lambda_j \otimes \Omega)\Gamma_j^{-1}\}, \end{aligned} \tag{30}$$

where

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

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

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

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

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

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

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

and $\tilde{\Lambda}_j$, $\tilde{\delta}_j$, $\hat{H}_j$, $\hat{\tau}_j$ are defined under Equation (6).

Suppose the current estimate of $\mathcal{H}$ is

$$\mathcal{H}^{(p)} = \left(\Omega^{(p)}, \Lambda_k^{(p)}, \left[\Lambda_j^{(p)}, \delta_j^{(p)}, \tau_{j0}{}^{(p)}\right], \quad j = 1, 2, \ldots, k-1\right). \tag{31}$$

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

(i) E-step: Compute the posterior expectations involved in (13), given data and $\mathcal{H}^{(p)}$.

(ii) M-step: Maximize $\mathcal{L}^*(\mathcal{H} \mid \mathcal{H}^{(p)})$ 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 Ω, maximize the following logarithmic function with respect to Ω,

$$\begin{aligned}
- & \left(\frac{\delta_k g_k + \sum_{j=1}^{k-1}(g_j^2 p + \delta_j g_j)}{2}\right) \log|\Omega^{-1}| - \frac{1}{2}\mathrm{tr}\{(\Lambda_k \otimes \Omega) E\left(\Sigma_{kk}^{-1} \mid D, \mathcal{H}^{(p)}\right)\} \\
- & \frac{1}{2}\sum_{j=1}^{k-1} \mathrm{tr}\left\{(\Lambda_j \otimes \Omega) E\left((\tau_j - \tau_{j0})\Gamma_j^{-1}(\tau_j - \tau_{j0})' \mid D, \mathcal{H}^{(p)}\right)\right\} - \frac{1}{2}\sum_{i=1}^{k-1} \mathrm{tr}(\Lambda_j \otimes \Omega) E(\Gamma_j^{-1} \mid D, \mathcal{H}^{(p)}) \\
- & \left(\frac{\delta_k g_k + \sum_{j=1}^{k-1}(g_j^2 p + \delta_j g_j)}{2}\right) \log|\Omega^{-1}| - \frac{1}{2}\left\{\mathrm{tr}(\Omega\hat{G}) + \sum_{j=1}^{k-1} \mathrm{tr}(\Omega\hat{K}_j)\right\},
\end{aligned}$$

where $\hat{G} = \sum_{i=1}^{g_k} G_{ii}$,

$$G = (G_{ij}) = (\Lambda_k \otimes I_p)\left\{E(\Sigma_{kk}^{-1}|D, \mathcal{H}^{(p)})\right\},$$

$\hat{K}_j = \sum_{i=1}^{g_k} K_{ii}$, and

$$K = (K_{ij}) = (\Lambda_j \otimes I_p)\left\{E(\Gamma_j^{-1}|D, \mathcal{H}^p)\right\} + E\{(\tau_j - \tau_{j0})\Gamma^{-1}(\tau_j - \tau_{j0})'|D, \mathcal{H}^{(p)}\}.$$

Then following Anderson (1984, Lemma 3.2.2), we obtain

$$\hat{\Omega}^{(p+1)} = \left(\frac{\delta_k g_k + \sum_{j=1}^{k-1}(g_j^2 p + \delta_j g_j)}{2}\right) \hat{P}^{-1},$$

where

$$\hat{P} = \hat{G} + \sum_{j=1}^{k-1} \hat{K}_j.$$

(b) To update the estimates of Λ_j, $\quad j = 1, 2, \ldots, k$, we consider two cases.

Case i: To update Λ_k maximize the following function:

$$
\begin{aligned}
- & \frac{\delta_k p}{2} \log |\Lambda_k^{-1}| - \frac{1}{2} \mathrm{tr}\{(\Lambda_k \otimes \Omega) E \left(\Sigma_{kk}^{-1} \mid D, \mathcal{H}\right)\} \\
- & \frac{\delta_k p}{2} \log |\Lambda_k^{-1}| - \frac{1}{2} \mathrm{tr}\{(\Lambda_k \tilde{L})
\end{aligned}
$$

This gives

$$\hat{\Lambda}_k^{(p+1)} = \delta_k p [\tilde{L}]^{-1},$$

where $\tilde{L} = (l_{ij})$, $l_{ij} = \mathrm{tr}(L_{ij})$ and

$$L = (I_{g_k} \otimes \Omega) E(\Sigma_{kk}^{-1} \mid D, \mathcal{H}^{(p)}).$$

Case ii: To update $\Lambda_j \quad j = 1, 2, \ldots, k-1$, maximize the following function:

$$-\sum_{j=1}^{k-1} \left(\frac{g_j p^2 + p\delta_j}{2}\right) \log |\Lambda_j^{-1}| - \frac{1}{2} \sum_{i=1}^{k-1} tr\left\{(\Lambda_j \otimes \Omega)[R_j + S_j]\right\}$$

where

$$R_j = E\left((\tau_j - \tau_{j0})\Gamma_j^{-1}(\tau_j - \tau_{j0})' \mid D, \mathcal{H}^{(p)}\right) \quad \text{and} \quad S_j = E(\Gamma_j^{-1} \mid D, \mathcal{H}^{(p)}).$$

The MLE of Λ_j is

$$\hat{\Lambda}_j^{(p+1)} = (g_j p^2 + p\delta_j)\tilde{M}^{-1}$$

where $\tilde{M} = (m_{ij})$, $m_{ij} = \mathrm{tr}(M)$ and $M = (I_{g_j} \otimes \Omega)(R_j + S_j)$.

(c) To upgrade the estimate of $\tau_{j0}, \quad j = 1, 2, \ldots, k-1$, maximize the following function

$$- \frac{1}{2} \sum_{j=1}^{k-1} tr\left\{(\Lambda_j \otimes \Omega) E\left((\tau_j - \tau_{j0})\Gamma_j^{-1}(\tau_j - \tau_{j0})' \mid D, \mathcal{H}^{(p)}\right)\right\}.$$

The MLE of $\boldsymbol{\tau}_{j0}$ we obtained is

$$\hat{\tau}_{j0}^{(p+1)} = \hat{W}_j^{(p)} \boldsymbol{\tau}_{j0}^{(p)} + (I - \hat{W}_j^{(p)})\hat{\boldsymbol{\tau}}_j,$$

where $\hat{W}_j^{(p)} = \left[(\Lambda_j^{(p)} \otimes \Omega^{(p)}) + (Y_{(j)}^{[g_j+1, \ldots, g_k]})'(Y_{(j)}^{[g_j+1, \ldots, g_k]})\right]^{-1}(\Lambda_j^{(p)} \otimes \Omega^{(p)})$ and $\hat{\boldsymbol{\tau}}_j = ((Y_{(j)}^{[g_j+1, \ldots, g_k]})'(Y_{(j)}^{[g_j+1, \ldots, g_k]}))^{-1}((Y_{(j)}^{[g_j+1, \ldots, g_k]})'Y^{[g_j^p]}$.

(d) To update the estimate of δ_j ($j = 1, 2, \ldots, k$), we maximize

$$
\begin{aligned}
- & \sum_{j=1}^{k} \frac{\delta_j g_j p}{2} \log 2 - \sum_{j=1}^{k-1} \sum_{i=1}^{g_j p} \log \Gamma\left(\frac{\delta_j - i + 1}{2}\right) + \sum_{j=1}^{k-1} \frac{\delta_j g_j}{2} \log |\Lambda_j \otimes \Omega| \\
+ & \frac{\delta_k}{2} \left\{ g_k \log 2 + \sum_{i=1}^{g_k} \Gamma\left(\frac{\hat{\delta}_k - i + 1}{2}\right) - \log |\hat{\Lambda}_k \otimes \hat{\Omega}| \right\} \\
- & \sum_{j=1}^{k-1} \frac{\delta_j}{2} \left\{ g_j \log 2 - \sum_{i=1}^{g_j} \Gamma\left(\frac{\hat{\delta}_k - i + 1}{2}\right) - \log |\hat{\Lambda}_j \otimes \hat{\Omega}| \right\}.
\end{aligned}
$$

Iterating these EM steps until convergence produces estimates for the hyperparameters including $(\Lambda_k, \Omega, [\Lambda_j, \tau_{0j}, \delta_j], j = 1, \ldots, k-1)$. These estimated hyperparameters can be used to form an estimate for $\Lambda^{[g]}$, associated with the spatial hypercovariance of the gauged sites, through the Barlett transformation. This estimate is relevant to the estimation of $\Lambda^{[u]}$ and $\tau_0^{[u]}$, the hyperparameters corresponding to the ungauged sites via the Sampson-Guttorp method as described in Section 4 and 5.