Multivariate Bayesian Spatial Interpolation using Gaussian-Generalized Inverted Wishart Model

B. M. Golam Kibria
The University of British Columbia [1]

Abstract

This paper has demonstrated the power and versatility of the Bayesian approach. It is one of a family of papers which developed a multivariate predictive methodology in the Bayesian framework. We assumed a Gaussian generalized inverted Wishart (GIW) model. First we derived the prediction distribution of the ungauged sites for known hyperparameters. The prediction distribution obtained, follow a matrix T distribution with appropriate covariance parameters and degrees of freedom. Then we developed an Expectation-Maximization (EM) algorithm to estimate the hyperparameters. Finally we obtain the predictive distribution for unobserved responses at ungauged sites. The results obtained in this paper will allow us to handle the data from different sites as well as multiple pollutants, and also where the observed data monitoring station follow a staircase structure.

Key words and Phrases: EM Algorithm; Inverted Wishart Distribution; Matrix T- Distribution; Posterior Distribution; Predictive Distribution; Spatial Interpolation.

1 Introduction

Risk assessments of air pollution often require estimates of the concentration levels at locations where there are no monitoring sites or at some time points yet to be observed 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

[1] Department of Statistics, The University of British Columbia, Vancouver, BC, V6T 1Z2, Email: gkibria@stat.ubc.ca

et al. (1995), Zidek (1997). 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.

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 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 measured 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 Bayesian spatial prediction has been considered by various researchers. To mention a few, Kitanidis (1986), Handcock et al (1993), Cui et al (1995), Gaudard et al (1999), and De Oliveira et al (1997).

In the paper, we develop a theoretical multivariate 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 (GIW) distribution. Specifically the responses are assumed to follow a Gaussian distribution and the corresponding covariance is assumed to follow a Generalized Inverted Wishart (GIW) prior distribution (Le et al (1999)). The developed methodology yields the joint predictive distribution for several locations and time points using all staircase observed data. The method is an multivariate extension for three blocks of the Bayesian developments discussed by Le et al (1999), to gain an interpolation theory for multiple pollutants, and therefore, enjoys all the corresponding advantages.

The structure of this paper is as follows. The main theoretical results are described in Section 2. The parameter estimation by the EM algorithm is discussed in Section 3. The predictive distribution for the unobserved responses at ungauged sites and the related parameters estimation are discussed in Section 4. Finally, concluding remarks have been added in Section 5.

2 Main Results

2.1 Notation

Throughout the paper, we adopt the standard notation used in Anderson (1984) when feasible. As well we let

$$\begin{aligned} n &= \text{number of time points (eg. number of days)}, \\ g &= g_1 + g_2 = \text{number of locations with monitors- called gauged sites}, \\ u &= \text{number of locations with no monitors - called ungauged sites, and} \\ b &= \text{number of pollutants.} \end{aligned}$$

The g sites are organized into 2 blocks of stations where responses at the g_1 stations in the first block are unobserved at n_1 time points and observed at $(n - n_1)$ time points. The responses at g_2 station in the second block are observed at $n(= n_1 + n_2)$ time points.

We represent the response variables at the ungauged and gauged sites by

$$\mathbf{Y}^{(c)} = [\mathbf{Y}^{(u)} | \mathbf{Y}^{(g)}] = \left[\mathbf{Y}^{(u)}, \begin{pmatrix} \mathbf{Y}_{11} & \mathbf{Y}_{12} \\ \mathbf{Y}_{21} & \mathbf{Y}_{22} \end{pmatrix} \right],$$

where $\mathbf{Y}^{(u)}$, a $n \times (u + b)$ matrix, denotes the responses at ungauged sites; $\mathbf{Y}^{(g)}$, a $n \times (g+b)$ matrix, denotes the responses at gauged sites; and also $\mathbf{Y}_{11}$, a $n_1 \times g_1 b$ matrix, denotes the unobserved responses at the g_1 gauged sites for the n_1 time points; $\mathbf{Y}_{21}$, a $(n - n_1) \times g_1 b$ matrix, denotes the responses at the g_1 gauged sites with observations; and $\mathbf{Y}_2 = (\mathbf{Y}_{12}, \mathbf{Y}_{22})$, $n \times g_2 b$ matrix, denotes the observed responses at the g_2 gauged sites;

In the first step, we developed the Bayesian prediction distribution for known hyperparameters. Here we adopt the same prior distribution as Le et al (1999) and present the multivariate interpolation for a Gaussian GIW prior. In the second step, we develop an EM algorithm to estimate the hyperparameters of the model.

2.2 The Gaussian GIW Model

We suppose $\mathbf{Y}^{(g)}$, be assumed to follows the Gaussian-Generalized-Inverted-Wishart model specified by:

$$\begin{cases} \mathbf{Y}^{(g)} \mid \mathbf{\Sigma} \sim N(\mathbf{O}, I_n \otimes \mathbf{\Sigma}); \\ \\ \mathbf{\Sigma} \sim GIW(\mathbf{\Lambda} \otimes \mathbf{\Omega}, \delta), \end{cases} \tag{1}$$

where $\mathbf{Y}^{(g)}$ is a $n \times gb$ response matrix, $\mathbf{\Sigma}$ is a $gb \times gb$ covariance matrix, $\mathbf{\Lambda}$ is a $b \times b$ covariance matrix of ions, which is assumed to be constant from site-to-site, $\otimes$ represents the Kronecker product between matrices and GIW represents the Generalized Inverted Wishart distribution developed by Brown, Le, Zidek (1994b). The distribution of $\mathbf{\Sigma}$ specified in terms of $\mathbf{\Sigma}_{22}$, $\mathbf{\Gamma}$ and $\boldsymbol{\tau}$, are as follows:

$$\begin{cases} \mathbf{\Sigma}_{22} \sim IW(\Lambda \otimes \mathbf{\Omega}_2, \delta_2); \\ \\ \mathbf{\Gamma} \sim IW(\mathbf{\Lambda} \otimes \mathbf{\Omega}_1, \delta_1); \\ \\ \boldsymbol{\tau}|\mathbf{\Gamma} \sim N\left(\boldsymbol{\tau}_0, \mathbf{H} \otimes \mathbf{\Gamma}\right), \end{cases} \tag{2}$$

where $\mathbf{\Sigma}_{22}$ is a $g_2 b \times g_2 b$ matrix, $\boldsymbol{\tau} = \mathbf{\Sigma}_{12}{\mathbf{\Sigma}_{22}}^{-1}$ is a $g_1 b \times g_2 b$ matrix, $\mathbf{\Gamma} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}{\mathbf{\Sigma}_{22}}^{-1}\mathbf{\Sigma}_{21}$ is a $g_1 b \times g_2 b$ matrix, $\mathbf{\Omega}_1$ is a $g_1 \times g_2$ matrix, $\mathbf{\Omega}_2$ is a $g_2 \times g_2$ matrix, $\mathbf{H} = (\mathbf{\Lambda} \otimes \mathbf{\Omega}_2)^{-1}$, and IW denotes the Inverted Wishart distribution. The hypercovariance matrix $(\mathbf{\Lambda} \otimes \mathbf{\Omega})$ has been partitioned comformably with of $\mathbf{\Sigma}$. The hyperparameters involved in the GIW model are written as

$$\mathcal{H} = \left\{\mathbf{\Lambda}, \mathbf{\Omega}_1, \mathbf{\Omega}_2, \delta_1, \delta_2, \boldsymbol{\tau}_0\right\}. \tag{3}$$

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).

2.3 The Posterior Distributions

Since the posterior distribution is an essential element of predictive distribution, we develop and investigate the relevant posterior distributions in this section.

Theorem 1. *The joint posterior density,* $f(\mathbf{\Sigma} \mid D, \mathcal{H})$ *is given by*

$$f(\mathbf{\Sigma} \mid D, \mathcal{H}) = f(\mathbf{\Sigma}_{22} \mid D, \mathcal{H}) f(\boldsymbol{\tau} \mid D, \mathbf{\Gamma}, \mathcal{H}) f(\mathbf{\Gamma} \mid D, \mathcal{H})$$

with

$$\mathbf{\Sigma}_{22} \mid D, \mathcal{H} \sim IW(\mathbf{\Lambda} \otimes \mathbf{\Omega}_2 + \mathbf{Y}_2'\mathbf{Y}_2, n + \delta_2),$$

$$\boldsymbol{\tau} \mid D, \mathbf{\Gamma}, \mathcal{H} \sim N\left(\boldsymbol{\tau}^*, (\mathbf{Y}_{22}'\mathbf{Y}_{22} + \mathbf{H}^{-1}) \otimes \mathbf{\Gamma}^{-1}\right) \tag{4}$$

$$\mathbf{\Gamma} \mid D, \mathcal{H} \sim IW(\mathbf{\Lambda} \otimes \mathbf{\Omega}_1 + \mathbf{Q}^*, n_2 + \delta_1),$$

where $\boldsymbol{\tau}^* = \mathbf{W}\boldsymbol{\tau}_0 + (\mathbf{I} - \mathbf{W})\tilde{\boldsymbol{\tau}}$, $\mathbf{W} = (\mathbf{Y}_{22}'\mathbf{Y}_{22} + \mathbf{H}^{-1})^{-1}\mathbf{H}^{-1}$, $\mathbf{Q}^* = (\mathbf{Y}_{21} - \mathbf{Y}_{22}\boldsymbol{\tau}_0)'[\mathbf{I}_{n_2} + \mathbf{Y}_{22}\mathbf{H}\mathbf{Y}_{22}']^{-1}(\mathbf{Y}_{21} - \mathbf{Y}_{22}\boldsymbol{\tau}_0)$.

Proof. The posterior distribution of $\mathbf{\Sigma}$ for given $data(D)$ and $\mathcal{H}$ is obatined as

$$\begin{aligned} f(\mathbf{\Sigma}|D, \mathcal{H}) &\propto f(\mathbf{Y}|\mathbf{\Sigma}) f(\mathbf{\Sigma}|\mathcal{H}) \\ &\propto f(\mathbf{Y}|\mathbf{\Sigma}) f(\mathbf{\Sigma}_{22}|\mathcal{H}) f(\boldsymbol{\tau}|\mathbf{\Gamma}, \mathcal{H}) f(\mathbf{\Gamma}|\mathcal{H}), \end{aligned} \tag{5}$$

where $f(\mathbf{Y}|\mathbf{\Sigma}) = f(\mathbf{Y}_{21}|\mathbf{Y}_2, \boldsymbol{\tau}, \mathbf{\Gamma}) f(\mathbf{Y}_2|\mathbf{\Sigma}_{22})$ and

$$\mathbf{Y}_{21}|D, \mathcal{H} \sim N(\mathbf{Y}_{22}\boldsymbol{\tau}, \mathbf{I}_{n_2} \otimes \mathbf{\Gamma}), \quad \text{and} \quad \mathbf{Y}_2|D, \mathcal{H} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_{\mathrm{n}_2} \otimes \mathbf{\Sigma}_{22}).$$

The prior distribution for $\mathbf{\Sigma}$ in terms of $\mathbf{\Sigma}_{22}$, $\boldsymbol{\tau}$ and $\mathbf{\Gamma}$ has been defined in (2). Using (2) and (5), the posteriror distribution is obtained as follows:

$$\begin{aligned} f(\mathbf{\Sigma}|D, \mathcal{H}) &\propto |\mathbf{\Gamma}|^{-\frac{n_2}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\mathbf{\Gamma}^{-1}(\mathbf{Y}_{21} - \mathbf{Y}_{22}\boldsymbol{\tau})'(\mathbf{Y}_{21} - \mathbf{Y}_{22}\boldsymbol{\tau})\right\} \\ &\times |\mathbf{\Sigma}_{22}|^{-\frac{n}{2}} \exp\Big\{ -\frac{1}{2}\mathrm{tr}\mathbf{\Sigma}_{22}^{-1}\mathbf{Y}_2'\mathbf{Y}_2\Big\} |\mathbf{\Lambda} \otimes \mathbf{\Omega}_2|^{\frac{\delta_2}{2}} |\mathbf{\Sigma}_{22}|^{\frac{\delta_2 + g_2 b + 1}{2}} \\ &\times \exp\Big\{ -\frac{1}{2}\mathrm{tr}\mathbf{\Sigma}_{22}^{-1}(\mathbf{\Lambda} \otimes \mathbf{\Omega}_2)\Big\} \; |\mathbf{\Lambda} \otimes \mathbf{\Omega}_1|^{\frac{\delta_1}{2}} |\mathbf{\Gamma}|^{\frac{\delta_1 + g_1 b + 1}{2}} \\ &\times \exp\Big\{ -\frac{1}{2}\mathrm{tr}\mathbf{\Gamma}^{-1}(\mathbf{\Lambda} \otimes \mathbf{\Omega}_1)\Big\} |\mathbf{H}|^{\frac{g_b}{2}} |\mathbf{\Gamma}|^{\frac{g_2 b}{2}} \\ &\times \exp\Big\{ -\frac{1}{2}\mathrm{tr}\mathbf{\Gamma}^{-1}((\boldsymbol{\tau} - \boldsymbol{\tau}_0)'\mathbf{H}^{-1}(\boldsymbol{\tau} - \boldsymbol{\tau}_0))\Big\}. \end{aligned}$$

After simplification the above equation, the posterior distribution of the parameters is obtained as

$$f(\mathbf{\Sigma}|D, \mathcal{H}) \propto |\mathbf{\Sigma}_{22}|^{-\frac{n + \delta_2 + g_2 b + 1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\Big[(\mathbf{\Lambda} \otimes \mathbf{\Omega}_2) + \mathbf{Y}_2'\mathbf{Y}_2\Big]\right\} |\mathbf{\Gamma}|^{-\frac{n_2 + \delta_1 + g_1 b + 1}{2}}$$

$$
\begin{aligned}
&\times \quad \exp\left\{\mathrm{tr}\boldsymbol{\Gamma}^{-1}\left\{(\boldsymbol{\Lambda}\otimes\boldsymbol{\Omega}_1) + (\mathbf{Y}_{21}-\mathbf{Y}_{22}\boldsymbol{\tau}_0)'[\mathbf{I}_{n_2}+\mathbf{Y}_{22}\mathbf{H}\mathbf{Y}_{22}']^{-1}\right\}\right\} \\
&\times \quad \{\{(\mathbf{Y}_{21}-\mathbf{Y}_{22}\boldsymbol{\tau}_0)\}\}\,|\boldsymbol{\Gamma}|^{-\frac{g_2 b}{2}}|\mathbf{Y}_{22}'\mathbf{Y}_{22}+\mathbf{H}^{-1}|^{-\frac{g_1 b}{2}} \\
&\times \quad \exp\left\{-\frac{1}{2}tr\left\{\boldsymbol{\Gamma}^{-1}(\boldsymbol{\tau}-\boldsymbol{\tau}^*)'[\mathbf{Y}_{22}'\mathbf{Y}_{22}+\mathbf{H}^{-1}]^{-1}(\boldsymbol{\tau}-\boldsymbol{\tau}^*)\right\}\right\}. \qquad (6)
\end{aligned}
$$

This proves the theorem.

The posterior means of $(\boldsymbol{\Sigma}_{22}, \tau, \boldsymbol{\Gamma})$ can be obtained directly from the Gaussian and Inverted Wishart distributions using the theorem 1. The predictive density for the unobserved responses at the gauged sites has developed in the following section.

2.4 The Predictive Distributions for Gauged Sites

This section provides the joint predictive distributions of all unobserved responses at gauged sites. 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.

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

$$
(\mathbf{Y}_{11} \mid D, \mathcal{H}) \sim t_{n_1\times g_1 b}(\mathbf{Y}_{12}\boldsymbol{\tau}^*, \mathbf{U}^*\otimes\mathbf{V}^{-1}, n_2 - g_1 b + \delta_1 - 1), \qquad (7)
$$

where

$$
\begin{aligned}
\mathbf{U}^* &= (n_2 - g_1 b + \delta_1 + 1)\times\mathbf{U}^{-1}, \\
\boldsymbol{\tau}^* &= \mathbf{W}\boldsymbol{\tau}_0 + (I-\mathbf{W})\tilde{\boldsymbol{\tau}}, \\
\mathbf{U} &= (\boldsymbol{\Lambda}\otimes\boldsymbol{\Omega}_1) + (\mathbf{Y}_{21}-\mathbf{Y}_{22}\boldsymbol{\tau}_0)'[\mathbf{I}_{n_2}+\mathbf{Y}_{22}\mathbf{H}\mathbf{Y}_{22}']^{-1}(\mathbf{Y}_{21}-\mathbf{Y}_{22}\boldsymbol{\tau}_0), \\
\mathbf{V} &= [\mathbf{I}_{n_1}+\mathbf{Y}_{12}(\mathbf{Y}_{22}'\mathbf{Y}_{22}+\mathbf{H}^{-1})^{-1}\mathbf{Y}_{12}'], \\
\mathbf{W} &= (\mathbf{Y}_{22}'\mathbf{Y}_{22}+\mathbf{H}^{-1})^{-1}\mathbf{H}^{-1}, \text{ and} \\
\tilde{\boldsymbol{\tau}} &= (\mathbf{Y}_{22}'\mathbf{Y}_{22})^{-1}\mathbf{Y}_{22}'\mathbf{Y}_{21}.
\end{aligned}
$$

Proof. The predictive distribution can be obtained as follows:

$$
p(\mathbf{Y}_{11}|D,\mathcal{H}) \propto \int_{\boldsymbol{\Gamma}>\mathbf{0}}\int_{\boldsymbol{\tau}} p(\mathbf{Y}_{11}|\boldsymbol{\tau},\boldsymbol{\Gamma},D)p(\boldsymbol{\tau},\boldsymbol{\Gamma}|D,\mathcal{H})d\boldsymbol{\Gamma}d\boldsymbol{\tau}, \qquad (8)
$$

where $\mathbf{Y}_{11}|\boldsymbol{\tau},\boldsymbol{\Gamma},D \sim N(\mathbf{Y}_{12}\boldsymbol{\tau},\ \mathbf{I}_{n1}\otimes\boldsymbol{\Gamma}^{-1})$, and the joint posterior distribution of $(\boldsymbol{\Gamma},\boldsymbol{\tau})$ is presented in Theorem 1. At the same time,

$$\begin{aligned} p(\mathbf{Y}_{11}|D,\mathcal{H}) &\propto \int_{\boldsymbol{\Gamma}>\mathbf{0}}\int_{\boldsymbol{\tau}} |\boldsymbol{\Gamma}|^{-\frac{n_1+n_2+\delta_1+gb+1}{2}}|\mathbf{U}|^{-\frac{n_2+\delta_1}{2}}|\mathbf{Y}_{22}'\mathbf{Y}_{22}+\boldsymbol{\Lambda}^{-1}\otimes\boldsymbol{\Omega}_2^{-1}|^{-\frac{g_1b}{2}} \\ &\times \exp\left\{-\frac{1}{2}\mathrm{tr}\boldsymbol{\Gamma}^{-1}\left\{(\mathbf{Y}_{11}-\mathbf{Y}_{12}\boldsymbol{\tau})'(\mathbf{Y}_{11}-\mathbf{Y}_{12}\boldsymbol{\tau})\right.\right. \\ &+ \left.\left.(\boldsymbol{\tau}-\boldsymbol{\tau}^*)'[\mathbf{Y}_{22}'\mathbf{Y}_{22}+\boldsymbol{\Lambda}^{-1}\otimes\boldsymbol{\Omega}_2^{-1}](\boldsymbol{\tau}-\boldsymbol{\tau}^*)+\mathbf{U}\right\}\right\} d\boldsymbol{\Gamma}d\boldsymbol{\tau}. \quad (9) \end{aligned}$$

The quadratic expression in (9) can be expressed as

$$(\mathbf{Y}_{11}-\mathbf{Y}_{12}\boldsymbol{\tau}^*)'\mathbf{V}(\mathbf{Y}_{11}-\mathbf{Y}_{12}\boldsymbol{\tau}^*)+(\boldsymbol{\tau}-\boldsymbol{\tau}^{(1)})'[\mathbf{Y}_{12}'\mathbf{Y}_{12}+(\mathbf{Y}_{22}'\mathbf{Y}_{22}'+\mathbf{H}^{-1})](\boldsymbol{\tau}-\boldsymbol{\tau}^{(1)}),$$

where $\boldsymbol{\tau}^{(1)}=\mathbf{W}^*\boldsymbol{\tau}^*+(\mathbf{I}_{g_2b}-\mathbf{W}^*)\tilde{\boldsymbol{\tau}}$, $\mathbf{W}^*=[\mathbf{Y}_{12}'\mathbf{Y}_{12}+(\mathbf{H}^*)^{-1}]^{-1}(\mathbf{H}^*)^{-1}$ and $\mathbf{H}^*=\mathbf{Y}_{22}'\mathbf{Y}_{22}+\mathbf{H}^{-1}$.

Equation (9), gives the prediction distribution of $\mathbf{Y}_{11}$ for given D and $\mathcal{H}$ as follows

$$p(\mathbf{Y}_{11}|D,\mathcal{H}) \propto |\mathbf{U}+(\mathbf{Y}_{11}-\mathbf{Y}_{22}\boldsymbol{\tau}^*)'\mathbf{V}(\mathbf{Y}_{11}-\mathbf{Y}_{22}\boldsymbol{\tau}^*)]|^{-\frac{n+\delta_1}{2}},$$

where $n+\delta_1$ can be written as

$$n+\delta_1=(n_1+g_1b-1)+(n_2-g_1b+\delta_1+1).$$

Finally, the predictive distribution of $\mathbf{Y}_{11}$ for given data D and $\mathcal{H}$, is obtained as follows:

$$p(\mathbf{Y}_{11}|D,\mathcal{H})=C_0\times|\mathbf{I}+\frac{1}{\delta^*}\mathbf{U}^*\times(\mathbf{Y}_{11}-\mathbf{Y}_{12}\boldsymbol{\tau}^*)'\mathbf{V}(\mathbf{Y}_{11}-\mathbf{Y}_{12}\boldsymbol{\tau}^*)]|^{-\frac{\delta^*+n_1+g_1b-1}{2}}, (10)$$

where C_0 is a normalizing constant, and $\delta^*=n_2-g_1b+\delta_1+1$. This proves the theorem.

For known hyperparameters, the unobserved response matrix $\mathbf{Y}_{11}$ follow a Matrix T distribution with covariance matrices $\mathbf{U}^{*-1}$ and $\mathbf{V}^{-1}$ and $(n_2-g_1b+\delta_1+1)$ degrees of freedom. However, for unknown hyperparameters, the prediction distribution can be approximate by its estimates obtained by using the EM algorithm.

3 Estimation of Hyperparameters

In this section, we discuss the estimation of the hyperparameter set $\mathcal{H}$ of equation (3). Specifically, we derive the EM algorithm corresponding to the model developed in the previous sections that our hyperparameters involve Kronecker structure, which has to be handled very carefully. The resulted estimators are the Type II maximum likelihood estimates. The EM approach has previously been discussed by various researchers. However, Dempster, et al (1977), Chen (1979), and very recenlty Le et al (1999) and Liu (1999) are notable among others.

We assume that the degrees of freedoms, δ_1, δ_2, follow a gamma prior distribution, where

$$\pi(\delta) \propto (\delta_1 \delta_2)^{\alpha-1} \exp\{-r(\delta_1 + \delta_2)\},$$

with α and r are specified.

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

$$\begin{aligned}\mathcal{L}(\mathcal{H} \mid \mathcal{H}^{(p)}) &= E\left(\log[f(Y, \mathbf{\Sigma} \mid \mathcal{H})\pi(\delta)] \mid D, \mathcal{H}^{(p)}\right) \\ &= E\left[\log f(Y|\mathbf{\Sigma}) \mid D, \mathcal{H}^{(p)}\right] + E\left[\log f(\mathbf{\Sigma} \mid \mathcal{H}) \mid D, \mathcal{H}^{(p)}\right] + \log \pi(\delta),\end{aligned} \tag{11}$$

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

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

$$\mathcal{L}^*(\mathcal{H} \mid \mathcal{H}^{(p)}) = E\left[\log f(\mathbf{\Sigma} \mid \mathcal{H}) \mid D, \mathcal{H}^{(p)}\right] + \log \pi(\delta) \tag{12}$$

at the E-step and maximize $\mathcal{L}^*$ over $\mathcal{H}$ at the M-step, where $f(\mathbf{\Sigma} \mid \mathcal{H})$ has given in (6).

With the parameterization introduced above

$$\log f(\mathbf{\Sigma}|\mathcal{H}) \quad \propto \quad \log c(\delta_1, \delta_2, g_1, g_2) + \log \pi(\delta) - \left(\frac{\delta_1 g_1 + \delta_2 g_2}{2}\right) \log |\mathbf{\Lambda}^{-1}| - \frac{b\delta_2}{2} |\mathbf{\Omega}_2^{-1}|$$

$$
\begin{aligned}
&- \quad \frac{b\delta_1}{2}|\boldsymbol{\Omega}_1^{-1}| \left(\frac{\delta_2+g_2b+1}{2}\right)\log|\boldsymbol{\Sigma}_{22}| - \left(\frac{\delta_1+g_1b+1}{2}\right)\log|\boldsymbol{\Gamma}| \\
&- \quad \frac{g_1 b}{2}\Big\{g_2\log|\boldsymbol{\Lambda}^{-1}| + b\log|\boldsymbol{\Omega}_2^{-1}|\Big\} - \frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{\Lambda}\otimes\boldsymbol{\Omega}_2)\right\} \\
&- \quad \frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Gamma}^{-1}(\boldsymbol{\Lambda}\otimes\boldsymbol{\Omega}_1)\right\} - \frac{g_2 b}{2}\log|\boldsymbol{\Gamma}| \\
&- \quad \frac{1}{2}\mathrm{tr}\Big[(\boldsymbol{\Lambda}\otimes\boldsymbol{\Omega}_2)(\boldsymbol{\tau}-\boldsymbol{\tau}_0)\boldsymbol{\Gamma}^{-1}(\boldsymbol{\tau}-\boldsymbol{\tau}_0)'\Big].
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\mathcal{L}^*(\mathcal{H}\mid\mathcal{H}^{(p)}) &= E\left[\log f(\boldsymbol{\Sigma}\mid\mathcal{H})\mid D,\mathcal{H}^{(p)}\right] + \log\pi(\delta) \\
&= Const + \log c(\delta_1,\delta_2,g_1,g_2) + \log\pi(\delta) - \left(\frac{\delta_1 g_1+\delta_2 g_2+g_1 g_2 b}{2}\right)\log|\boldsymbol{\Lambda}^{-1}| \\
&- \left(\frac{b\delta_2+g_1b^2}{2}\right)|\boldsymbol{\Omega}_2^{-1}| - \frac{b\delta_1}{2}|\boldsymbol{\Omega}_1^{-1}| - \left(\frac{\delta_2+g_2b+1}{2}\right)E\left[\log|\boldsymbol{\Sigma}_{22}|\mid D,\mathcal{H}^{(p)}\right] \\
&- \left(\frac{\delta_1+gb+1}{2}\right)E\left[\log|\boldsymbol{\Gamma}|\mid D,\mathcal{H}^{(p)}\right] - \frac{1}{2}tr\Big[E\left[\boldsymbol{\Sigma}_{22}^{-1}\mid D,\mathcal{H}^{(p)}\right](\boldsymbol{\Lambda}\otimes\boldsymbol{\Omega}_2)\Big] \\
&- tr\Big[E\left[\boldsymbol{\Gamma}^{-1}\mid D,\mathcal{H}^{(p)}\right](\boldsymbol{\Lambda}\otimes\boldsymbol{\Omega}_1)\Big] - \frac{1}{2}tr\Big[E\Big((\boldsymbol{\tau}-\boldsymbol{\tau}_0)\boldsymbol{\Gamma}^{-1}(\boldsymbol{\tau}-\boldsymbol{\tau}_0)'|D,\mathcal{H}^{(p)}\Big) \\
&\times (\boldsymbol{\Lambda}\otimes\boldsymbol{\Omega}_2)\Big], \qquad (13)
\end{aligned}
$$

where

$$
\begin{aligned}
E(\log|\boldsymbol{\Sigma}_{22}|\mid D,\mathcal{H}^{(p)}) &= -g_2 b\log 2 - \sum_{i=1}^{g_2 b}\eta\left(\frac{n+\delta_2^{(p)}-i+1}{2}\right) \\
&+ \log\left\{|\boldsymbol{\Lambda}^{(p)}\otimes\boldsymbol{\Omega}_2^{(p)} + \mathbf{Y}_2'\mathbf{Y}_2|\right\}, \\
E(\log|\boldsymbol{\Gamma}|\mid D,\mathcal{H}^{(p)}) &= -g_1 b\log 2 - \sum_{i=1}^{g_1 b}\eta\left(\frac{n_2+\delta_1^{(p)}-i+1}{2}\right) + \log\Big\{|\boldsymbol{\Lambda}^{(p)}\otimes\boldsymbol{\Omega}_1^{(p)} \\
&+ (\mathbf{Y}_{21}-\mathbf{Y}_{22}\boldsymbol{\tau}_0^{(p)})'\Big[\mathbf{I}_{n_2}+\mathbf{Y}_{22}(\boldsymbol{\Lambda}^{(p)}\otimes\boldsymbol{\Omega}_1^{(p)})^{-1}\mathbf{Y}_{22}\Big](\mathbf{Y}_{21}-\mathbf{Y}_{22}\boldsymbol{\tau}_0^{(p)})|\Big\}, \\
E(\boldsymbol{\Sigma}_{22}^{-1}\mid D,\mathcal{H}^{(p)}) &= (n+\delta_2^{(p)})[(\boldsymbol{\Lambda}^{(p)}\otimes\boldsymbol{\Omega}_1^{(p)}) + \mathbf{Y}_2'\mathbf{Y}_2]^{-1}, \\
E(\boldsymbol{\Gamma}^{-1}\mid D,\mathcal{H}^{(p)}) &= (n_2+\delta_1^{(p)})\Big[(\boldsymbol{\Lambda}^{(p)}\otimes\boldsymbol{\Omega}_1^{(p)}) + (\mathbf{Y}_{21}-\mathbf{Y}_{22}\boldsymbol{\tau}_0^{(p)})' \\
&\times \Big[\mathbf{I}_{n_2}+\mathbf{Y}_{22}(\boldsymbol{\Lambda}^{(p)}\otimes\boldsymbol{\Omega}_1^{(2)})^{-1}\mathbf{Y}_{22}\Big](\mathbf{Y}_{21}-\mathbf{Y}_{22}\boldsymbol{\tau}_0^{(p)})\Big]^{-1}, \\
E(\boldsymbol{\tau}\boldsymbol{\Gamma}^{-1}\boldsymbol{\tau}'\mid D,\mathcal{H}^{(p)}) &= \delta_1^{(p)}\boldsymbol{\tau}^{(p)}(\boldsymbol{\Lambda}^{(p)}\otimes{\boldsymbol{\Omega}_1}^{(p)})^{-1}\boldsymbol{\tau}^{(p)'} + g_1(\boldsymbol{\Lambda}^{(p)}\otimes\boldsymbol{\Omega}_p^{(p)})^{-1}, \text{ and} \\
E(\boldsymbol{\tau}\boldsymbol{\Gamma}^{-1}\mid D,\mathcal{H}^{(p)}) &= \delta_1^{(p)}\boldsymbol{\tau}^{(p)}(\boldsymbol{\Lambda}^{(p)}\otimes{\boldsymbol{\Omega}_1}^{(p)})^{-1}\boldsymbol{\tau}^{(p)'}.
\end{aligned}
$$

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

$$
\mathcal{H}^{(p)} = \left(\boldsymbol{\Lambda}^{(p)},\boldsymbol{\Omega}_1^{(p)},\boldsymbol{\Omega}_2^{(p)},\delta_1^{(p)},\delta_2^{(p)},{\boldsymbol{\tau}_0}^{(p)}\right). \qquad (14)
$$

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 $\mathbf{\Lambda}$, maximize the following logarithmic function with respect to $\mathbf{\Lambda}$,

$$-\left(\frac{\delta_1 g_1 + \delta_2 g_2 + g_1 g_2 b}{2}\right) \log |\mathbf{\Lambda}^{-1}| - \frac{1}{2} tr\{\mathbf{\Lambda}\tilde{\mathbf{P}}\},$$

where

$$\tilde{\mathbf{P}} = \tilde{\mathbf{G}} + \tilde{\mathbf{K}},$$

also $\tilde{\mathbf{G}} = (g_{ij})$, $g_{ij} = tr(\mathbf{G}_{ij})$, and

$$\mathbf{G} = (\mathbf{G}_{ij}) = (\mathbf{I}_b \times \mathbf{\Omega}_2) \Big\{ E(\mathbf{\Sigma}_{22}^{-1} | D, \mathcal{H}^{(p)}) + E\{(\boldsymbol{\tau} - \boldsymbol{\tau}_0)\mathbf{\Gamma}^{-1}(\boldsymbol{\tau} - \boldsymbol{\tau}_0)' | D, \mathcal{H}^{(p)}\} \Big\}$$

and $\tilde{\mathbf{K}} = (k_{ij})$, $k_{ij} = tr(\mathbf{K}_{ij})$, and

$$\mathbf{K} = (\mathbf{K}_{ij}) = (\mathbf{I}_b \times \mathbf{\Omega}_1) \Big\{ E(\mathbf{\Gamma}^{-1} | D, \mathcal{H}^p) \Big\}$$

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

$$\tilde{\mathbf{\Lambda}}^{(p+1)} = (\delta_1 g_1 + \delta_2 g_2 + g_1 g_2 d)\ \tilde{\mathbf{P}}^{-1}.$$

(b) To update the estimates of $\mathbf{\Omega}_1$, maximize

$$-\frac{\delta_1 b}{2} \log |\mathbf{\Omega}_1^{-1}| - \frac{1}{2} tr(\mathbf{\Omega}_1 \tilde{\mathbf{L}}),$$

where $\tilde{\mathbf{L}} = \mathbf{L}_{11} + \mathbf{L}_{22} +, \ldots, \mathbf{L}_{bb}$ and $\mathbf{L} = (\mathbf{L}_{ij}) = E\Big[\mathbf{\Gamma}^{-1} | D, \mathcal{H}^p\Big](\mathbf{\Lambda} \otimes \mathbf{I}_{g_1})$. The maximum likelihood estimate of $\mathbf{\Omega}_1$ is obtained as

$$\tilde{\mathbf{\Omega}}_1^{(p+1)} = b\delta_1 \tilde{\mathbf{L}}^{-1}.$$

(c) To update the estimates of $\mathbf{\Omega}_2$, maximize

$$-\left(\frac{b\delta_2 + g_1 b^2}{2}\right) \log |\mathbf{\Omega}_2^{-1}| - \frac{1}{2} tr(\mathbf{\Omega}_2 \tilde{\mathbf{H}}),$$

where $\tilde{\mathbf{H}} = \mathbf{H}_{11} + \mathbf{H}_{22} + \ldots + \mathbf{H}_{bb}$ and $\mathbf{H} = (\mathbf{H}_{ij}) = E\Big[\mathbf{\Sigma}_{22}^{-1}|D, \mathcal{H}^{(p)}\Big](\mathbf{\Lambda} \otimes \mathbf{I}_{g_2})$. Then the maximum likelihood estimate of $\mathbf{\Omega}_2$ is obtained as

$$\tilde{\mathbf{\Omega}}_2^{(p+1)} = (b\delta_2 + g_1 b)\tilde{\mathbf{H}}^{-1}.$$

(d) To upgrade the estimate of $\boldsymbol{\tau}$, maximize the following function

$$\begin{aligned}
- & \frac{1}{2}\Big[(n_2 + \delta_1)(\boldsymbol{\tau}^* - \boldsymbol{\tau}_0)\Big[(\mathbf{\Lambda}^{(p)} \otimes \mathbf{\Omega}_1^{(p)}) + (\mathbf{Y}_{21} - \mathbf{Y}_{22}\boldsymbol{\tau}_0)' \\
\times & [\mathbf{I}_{n_2} + \mathbf{Y}_{22}(\mathbf{\Lambda}^{(p)} \otimes \mathbf{\Omega}_2^{(p)})^{-1}\mathbf{Y}_{22}'](\mathbf{Y}_{21} - \mathbf{Y}_{22}\boldsymbol{\tau}_0)\Big](\boldsymbol{\tau}^* - \boldsymbol{\tau}_0)' \\
+ & g_1[\mathbf{Y}_{22}'\mathbf{Y}_{22} + (\mathbf{\Lambda}^{(p)} \otimes \mathbf{\Omega}_2^{(p)})]\Big].
\end{aligned}$$

The MLE of $\boldsymbol{\tau}_0$ is obtained as

$$\tilde{\boldsymbol{\tau}}_0^{(p+1)} = \tilde{\mathbf{W}}^{(p)}\boldsymbol{\tau}_0^{(p)} + (\mathbf{I} - \tilde{\mathbf{W}}^{(p)})\tilde{\boldsymbol{\tau}},$$

where $\tilde{\mathbf{W}}^{(p)} = \Big[\mathbf{Y}_{22}'\mathbf{Y}_{22} + (\mathbf{\Lambda}^{(p)} \otimes \mathbf{\Omega}_2^{(p)})\Big]^{-1}(\mathbf{\Lambda}^{(p)} \otimes \mathbf{\Omega}_2^{(p)})$ and $\tilde{\boldsymbol{\tau}} = (\mathbf{Y}_{22}'\mathbf{Y}_{22})^{-1}\mathbf{Y}_{22}'\mathbf{Y}_{21}$.

(e) To update the estimates of δ, we maximize the following function:

$$\begin{aligned}
- & \left(\frac{b(\delta_1 g_1 + \delta_2 g_2)}{2}\right)\log 2 - \sum_{i=1}^{g_j b}\log\mathbf{\Gamma}\left(\frac{\delta_j - i + 1}{2}\right) + (\alpha - 1)(\log\delta_1 + \log\delta_2) \\
- & (\delta_1 + \delta_2)r + \sum_{j=1}^{2}\delta_j\log|\mathbf{\Lambda} \otimes \mathbf{\Omega}_j| + \left(\frac{\delta_2 + g_1 b + 1}{2}\right)\left\{g_2\ b\ \log 2 + \sum_{i=1}^{g_2 b}\eta\left(\frac{n + \delta_2^{(p)} - i + 1}{2}\right)\right. \\
- & \log\Big\{|(\mathbf{\Lambda}^{(p)} \otimes \mathbf{\Omega}_2^{(p)}) + \mathbf{Y}_{22}'\mathbf{Y}_{22}|\Big\}\Big\} + \frac{\delta_1 + gb + 1}{2}\left\{g_1 b\log 2 + \sum_{i=1}^{g_2 b}\eta\left(\frac{n_2 + \delta_1^{(p)} - i + 1}{2}\right)\right. \\
- & \log\Big\{|(\mathbf{\Lambda}^{(p)} \otimes \mathbf{\Omega}_1^{(p)}) + (\mathbf{Y}_{21} - \mathbf{Y}_{22}\boldsymbol{\tau}_0^{(p)})' \\
\times & [\mathbf{I}_{n_2} + \mathbf{Y}_{22}(\mathbf{\Lambda}^{(p)} \otimes \mathbf{\Omega}_2^{(p)})^{-1}\mathbf{Y}_{22}'](\mathbf{Y}_{21} - \mathbf{Y}_{22}\boldsymbol{\tau}_0^{(p)})|\Big\}\Big\}
\end{aligned}$$

It follows that the estimates of δ_j, $\mathbf{\Omega}_j$,$(j = 1, 2)$ and $\mathbf{\Lambda}$ can iteratively be updated by solving the following equation

$$\begin{aligned}
- & \sum_{i=1}^{g_j b}\eta\left(\frac{\delta_j^{(p+1)} - i + 1}{2}\right) + \frac{\alpha - 1}{\delta_j^{(p+1)}} - r + \log|\mathbf{\Lambda}^{(p+1)} \otimes \mathbf{\Omega}_j^{(p+1)}| + \sum_{i=1}^{g_2 b}\eta\left(\frac{n + \delta_2^{(p)} - i + 1}{2}\right) \\
- & \log\Big\{|(\mathbf{\Lambda}^{(p)} \otimes \mathbf{\Omega}_2^{(p)}) + \mathbf{Y}_{22}'\mathbf{Y}_{22}|\Big\} + \sum_{i=1}^{g_2 b}\eta\left(\frac{n_2 + \delta_1^{(p)} - i + 1}{2}\right) \\
- & \log\Big\{|(\mathbf{\Lambda}^{(p)} \otimes \mathbf{\Omega}_1^{(p)}) + (\mathbf{Y}_{21} - \mathbf{Y}_{22}\boldsymbol{\tau}_0^{(p)})'[\mathbf{I}_{n_2} + \mathbf{Y}_{22}(\mathbf{\Lambda}^{(p)} \otimes \mathbf{\Omega}_2^{(p)})^{-1}\mathbf{Y}_{22}'](\mathbf{Y}_{21} - \mathbf{Y}_{22}\boldsymbol{\tau}_0^{(p)})|\Big\}
\end{aligned}$$

Iterating these EM steps until convergence produces estimates $\tilde{\mathcal{H}}$ for the hyperparameters $\mathcal{H} = (\mathbf{\Lambda}, \mathbf{\Omega}_1, \mathbf{\Omega}_2, \delta_1, \delta_2, \boldsymbol{\tau}_o)$.

4 Predictive Distributions for Ungauged Sites

This section provides the joint predictive distributions of all unobserved responses at ungauged sites.

Let $\mathbf{Y}^{(\mathbf{c})} = \left[\mathbf{Y}^{(u)}|\mathbf{Y}^{(g)}\right]$, be assumed to follow the Gaussian- Inverted-Wishart model specified by:

$$\begin{cases} \mathbf{Y}^{(c)} \mid \tilde{\mathbf{\Sigma}} \sim N(\mathbf{O}, \mathbf{I}_n \otimes \tilde{\mathbf{\Sigma}}); \\ \\ \tilde{\mathbf{\Sigma}} \sim IW(\mathbf{\Lambda} \otimes \tilde{\mathbf{\Omega}}, \delta^{(u)}), \end{cases} \tag{15}$$

where $\mathbf{Y}^{(\mathbf{c})}$ is a $n \times (u+g)b$ response matrix, $\tilde{\mathbf{\Sigma}}$ is a $(u+g)b \times (u+g)b$ covariance matrix, $\mathbf{\Lambda}$ is a $b \times b$ covariance matrix of ions, which is assumed to be constant from site-to-site, and

$$\tilde{\mathbf{\Omega}} = \begin{pmatrix} \tilde{\mathbf{\Omega}}_{uu} & \tilde{\mathbf{\Omega}}_{ug} \\ \tilde{\mathbf{\Omega}}_{gu} & \tilde{\mathbf{\Omega}}_{gg} \end{pmatrix}$$

is a $(u+g) \times (u+g)$ matrix. Then following Le et al (1999), we obtain the posterior distribution of $\tilde{\mathbf{\Sigma}}$ for given $\mathbf{Y}^{(c)}$ as

$$\tilde{\mathbf{\Sigma}} \mid \mathbf{Y}^{(c)} \sim IW(\mathbf{\Lambda} \otimes \tilde{\mathbf{\Omega}} + \mathbf{Y}^{(c)\prime}\mathbf{Y}^{(c)}, \delta^{(u)} + n), \tag{16}$$

and the marginal distribution of $\mathbf{Y}^{(c)}$ as

$$\mathbf{Y}^{(c)} \sim t_{n\times(u+g)}(0, (\delta^{(u)} - (g+u) + 1)^{-1} I_n \otimes \mathbf{\Lambda} \otimes \tilde{\mathbf{\Omega}}, \delta^{(u)} - (g+u) + 1). \tag{17}$$

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

$$\begin{aligned} (\mathbf{Y}^{(u)} \mid \mathbf{Y}^{(g)}, \mathcal{H}^*) \;\; \sim \;\; & t_{n\times ub}\Big\{\mathbf{Y}^{(g)}\mathbf{I}_b \otimes \boldsymbol{\tau}_0^{(u)}, (\delta^{(u)} - (u+g) + 1)^{-1} \\ \times \;\; & (\mathbf{I}_n + \mathbf{Y}^{(g)}\mathbf{Z}\mathbf{Y}^{(g)\prime}) \otimes \mathbf{\Omega}^{(u,g)}, \delta^{(u)} - (u+g) + 1\Big\}, \end{aligned} \tag{18}$$

where $\mathbf{Y}^{(g)}$ matrix contains both observed and interpolated responses, $\mathcal{H}^* = (\tilde{\mathbf{\Omega}}, \lambda, \delta^{(u)})$, $\boldsymbol{\tau}_0^{(u)} = \tilde{\mathbf{\Omega}}_{gg}^{-1}\tilde{\mathbf{\Omega}}_{gu}$, $\mathbf{Z} = \mathbf{\Lambda}^{-1} \otimes \tilde{\mathbf{\Omega}}_{gg}^{-1}$ and $\mathbf{\Omega}^{(u,g)} = \frac{\delta^{(u)}-(u+g)+1}{\delta^{(u)}-u+1}[\tilde{\mathbf{\Omega}}_{uu} - \tilde{\mathbf{\Omega}}_{ug}\tilde{\mathbf{\Omega}}_{gg}^{-1}\tilde{\mathbf{\Omega}}_{gu}]$.

Proof: It follows from (17) and the property of conditional distribution of $\mathbf{Y}^{(u)}$ for gievn $\mathbf{Y}^{(g)}$ of a matrix-T distribution.

4.1 Parameters Estimation

The hyperparameters $\mathbf{\Lambda}$, $\tilde{\mathbf{\Omega}}_{gg}(=\mathbf{\Omega})$ have already been estimated in section 3 by EM algorithm. The remaining hyperparameters, $\tilde{\mathbf{\Omega}}_{gu}$, $\tilde{\mathbf{\Omega}}_{uu}$, and $\tilde{\mathbf{\Omega}}_{ug}$ are estimable following Le et al (1999), by using the Sampson-Guttorp method. It is noted that SG method is designed to extend the spatial covariance from the gauged to the ungauged sites. Therefore, this method is used to extend $\tilde{\mathbf{\Omega}}_{gg} = \tilde{\mathbf{\Omega}}$ to estimate $\tilde{\mathbf{\Omega}}_{gu}$, and $\tilde{\mathbf{\Omega}}_{uu}$, and $\tilde{\mathbf{\Omega}}_{ug}$.

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

The degrees of freedom $\delta^{(u)}$, have to selected before interpolate the unobserved data. The best guesses for the selection are to be $\delta^{(u)}$ as

$$\delta^{(u)} = \max(\tilde{\delta}_1, \tilde{\delta}_2) \quad \text{or} \quad \frac{\tilde{\delta}_1 + \tilde{\delta}_2}{2},$$

subject to condition that $\delta^{(u)} \geq ub$.

5 Concluding Remarks

This paper has developed a Bayesian approach for multivariate spatial and temporal interpolation problem. This approach is an extention of the Bayesian methodology for spatial interpolation developed by Le, Sun and Zidek (1999) to gain an interpolation theory for multiple pollutants. We assumed 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 prediction distribution obtained, follow a matrix T distribution with appropriate covariance parameters and degrees of freedom. We also developed an EM algorithm which is a significant contribution to estimate the unknown hyperparameters for multiple pollutants model. We also developed a predictive distribution for the unobserved responses at ungauged sites. The results obtained in this paper will allow us to analyses the data from different sites as well as multiple pollutants, where the observed data monitoring station follow a staircase structure.

For the brevity of the paper, we have presented three blocks case. The development of this approach for general case (k blocks, which is straight forward to this one) and the practical application of this method is currently under investigation.

ACKNOWLEDGEMENTS

I am indebted to Drs. Nhu Le, Li Sun and James V. Zidek for introducing me to this problems as well for their encouragement and support. Special thanks to Dr. Li Sun for his guidance in the preperation of this paper. I gratefully acknowledge the financial support from grants hold by Professor James V. Zidek.

REFERENCES

Anderson, T. W. (1984). An Introduction To Multivariate Statistical Analysis. Second Edition, Wiley.

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

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

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. Wiely.

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.

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.

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. (1999). Bayesian spatial interpolation and backcasting using Gaussian-generalized interted wishart Model. TR # 185, Dept. of Statistics, The University of British Columbia.

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

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: Wiely 251-268.