The UNIVERSITY OF BRITISH
COLUMBIA
DEPARTMENT OF STATISTICS
TECHNICAL REPORT # 249

TEMPORAL PREDICTION WITH
A BAYESIAN SPATIAL
PREDICTOR: AN APPLICATION
TO OZONE FIELDS

BY

YIPING DOU
NHU D LE
JAMES V ZIDEK

JUNE 2009

Temporal prediction with a Bayesian spatial predictor: an application to ozone fields. *

Yiping Dou, Nhu D Le and James V Zidek

June 5, 2009

Abstract

This report shows how a Bayesian hierarchical approach we call Bayesian spatial prediction (BSP), originally developed for spatial prediction, can be used to temporally predict (i.e. forecast) a future multivariate response vector at a site which monitors a space–time field. The theory needed for that approach, which uses available past data from all such sites, is presented and demonstrated by application to Chicago AQS ozone fields. Moreover it is compared with two other approaches: dynamic linear modeling, a common, state–space approach to forecasting; NAIVE, a naive approach that relies only on common weekly and hourly effects. Overall the results shows BSP to be the best of these approaches for the data considered.

Keywords: Dynamic linear model, hierarchical Bayes, ozone, space–time fields, Bayesian spatial prediction after pre–filtering.

1   Introduction

This report shows how a Bayesian hierarchical approach we call Bayesian spatial prediction (BSP), originally developed for spatial prediction, can be used to temporally predict (i.e. forecast) a future multivariate response vector at a site which monitors a space–time field. The theory needed for that approach, which uses available past data from all such sites, is presented and demonstrated by application to Chicago AQS ozone fields.

Concerns about ozone go back at least 100 years since Dr. Henry Antoine Des Voeux mentioned smog, of which ozone is a constituent, in a paper, "Fog and Smoke", that he presented in 1905. However much added impetuous came from the US Air Clean Act of 1970 that led to the designation of six primary air pollutants (carbon monoxide, lead, nitrogen dioxide, ozone, particulate matter and sulfur dioxide) as criteria pollutants subject to regulation. These are the pollutants deemed to be risks to human health (for which primary standards are set) and human welfare (secondary standards).

US EPA (Environmental Protection Agency) monitors ground–level ozone over the entire US. The monitoring stations are irregularly distributed over the country and provide in particular hourly measurements of ground–level ozone concentrations, the focus of this report. More specifically it concerns one day ahead forecasts of those concentrations based on current day (and previous) concentrations, forecasts that are commonly provided nowadays in urban areas to alert individuals in groups susceptible to adverse health outcomes. In fact, the forecaster may need to answer questions such as: "What will the ozone concentration level be at 2 p.m. tomorrow given all data until 10 a.m. today?"; or, "What will the ozone levels be tomorrow if I have all the measurements up to today?"

To model the space–time fields we use the Bayesian spatial prediction after pre–filtering (BSP) approach, also called the Bayesian hierarchical kriging method (Le and Zidek 1992, 2006; Brown et al. 1994; Le et al. 1997; Zidek et al. 2002). The multivariate BSP approach can be adapted to answer questions like those above, by creating a 24–dimensional multivariate response variable with the daily 24 hourly univariate responses, treated formally as if they were 24 "species" or "pollutants"; each entry therein represents one measurement for each of the successive 24 hours. The multivariate model exploits the strong dependence in the sequence of hourly responses and allows the as yet unobserved, say 24_{th} hour response to borrow strength from its predecessors. To meet the BSP's assumption of independence, we use two subsequences of the 24–dimensional response vectors. More precisely we create two subdata matrices using the 24–dimensional vectors, the first from the odd days and a second, the even days. Each subdata matrix is then amenable to application of available software for the BSP (*EnviRo.stat.ubc.ca*; Dou et al. 2007, 2008). The resulting odd–even pairs of hyperparameter estimates from the software can be averaged to form "estimates" of hyperparameters based on all the data. Finally, one–day–ahead forecasts at gauged sites given those estimates of hyperparameters and observed responses, can be obtained, with say 95% forecast intervals, from the posterior distribution. However alternative methods for constructing those forecasts are available.

In fact, Section 2 presents three methodologies to forecast future responses at gauged sites: the multivariate BSP, DLM and NAIVE. Section 2.1 shows how to construct the above odd–day and even–day sequences and the corresponding subdata matrices along with the multivariate settings of the BSP model to predict each one

2

of the 24 responses next day. Moreover, their predictive posterior distributions are developed and the corresponding pointwise predictive intervals at each gauged site, constructed. Section 2.2 extends the results in Section 2.1 to predict r–step–ahead responses for any $r \in \mathcal{N}$. Section 2.3 illustrates the r–step–ahead prediction by the DLM approach. Section 2.4 presents the r–step–ahead temporal prediction by NAIVE approach. Section 3 implements the multivariate BSP forecaster for the Chicago AQS ozone database (2000). Section 4 presents the results and comparisons of the one–day–ahead prediction by the three approaches at gauged sites. Section 5 summarizes the success of the newly developed temporal prediction by the BSP and potential problems of this method.

2   Methodology

This section introduces the multivariate BSP, DLM and NAIVE approaches to forecasting future responses. We start from the one–day–ahead prediction using the BSP in Section 2.1, and then extends those results to r–step–ahead ($r \in \mathcal{N}$) prediction in the following subsection. Section 2.3 describes the DLM method for forecasting. Moreover, Section 2.4 illustrates the NAIVE approach and compares its forecasts with those from the BSP and DLM approaches.

2.1   One–day–ahead prediction with the BSP

Suppose $Y_{t,i}^{[g_j^m]}$ represents the unobserved i^{th} response variable at time point t, monitoring ("gauged") site j, and $Y_{t,i}^{[g_j^o]}$, the observed response variable, for $t = 1, \ldots, n$, $i = 1, \ldots, p$, and $j = 1, \ldots, g$. We have 121 days of hourly observed ozone over 14 gauged sites. To assess the BSP forecast, we set aside the observations for the last day, Day 121. We then predict the ozone concentration levels on that day at gauged sites, given the hourly observations from day 1 to day 120. To exploit the dependence among the hourly responses, we use a multivariate model with each day's hourly ozone concentrations in a single 24 dimensional response vector. Hence, we have $n = 120$ days, $p = 24$ responses and $g = 14$ gauged sites. For our demonstration, we make forecasts in two arbitrarily selected cases, (i) 11 P.M. and (ii) any single hour during the period from 0 A.M. to 10 P.M. on Day 121.

- **Case 1: Predict the concentration during the last hour (i.e., 11 P.M.) of the 121st day.**

  Call the corresponding multivariate BSP model "Model–1". One of the sub-data matrices, i.e., "odd–day–response", can be formed by $\{\mathbf{U}_t^{(1)} : 1 \times gp, \quad t = 1, \ldots, 60\}$, where $\mathbf{U}_t^{(1)} = (Y_{2t-1,1}^{[g_1^o]}, \ldots, Y_{2t-1,p}^{[g_g^o]})$; the other, "even–day–response", by $\{\mathbf{V}_t^{(1)} : 1 \times gp, \quad t = 1, \ldots, 59\}$, where $\mathbf{V}_t^{(1)} = (Y_{2t,1}^{[g_1^o]}, \ldots, Y_{2t,p}^{[g_g^o]})$. In other

3

words, the observed response variables from the 1$^{\text{st}}$ day to the 119$^{\text{th}}$ day are used as the "data". After that, each of the subdata matrices is used with Model–1 to find estimates of the hyperparameters. Moreover, "approximate" estimates of the hyperparameters given all the observed response variables at gauged sites are obtained by averaging each pair of the hyperparameters estimated by the BSP using these two matrices respectively.

- **Case 2: Predict the response variable at the $(k-1)^{\text{th}}$ hour of the 121$^{\text{st}}$ day, for $k = 2, \ldots, p$.**

  Call the corresponding multivariate BSP model "Model–k". One of the subdata matrices can be formed by $\{\mathbf{U}_t^{(k)} : 1 \times gp, \quad t = 1, \ldots, 59\}$, where $\mathbf{U}_t^{(k)} = (Y_{2t-1,k}^{[g_1^o]}, \ldots, Y_{2t-1,p}^{[g_1^o]}, Y_{2t,1}^{[g_1^o]}, \ldots, Y_{2t,k-1}^{[g_1^o]}, \ldots, Y_{2t-1,k}^{[g_g^o]}, \ldots, Y_{2t-1,p}^{[g_g^o]}, Y_{2t,1}^{[g_g^o]}, \ldots, Y_{2t,k-1}^{[g_g^o]})$; the other by $\{\mathbf{V}_t^{(k)} : 1 \times gp : \quad t = 1, \ldots, 59\}$, where $\mathbf{V}_t^{(k)} = (Y_{2t,k}^{[g_1^o]}, \ldots, Y_{2t,p}^{[g_1^o]}, Y_{2t+1,1}^{[g_1^o]}, \ldots, Y_{2t+1,k-1}^{[g_1^o]}, \ldots, Y_{2t,k}^{[g_g^o]}, \ldots, Y_{2t,p}^{[g_g^o]}, Y_{2t+1,1}^{[g_g^o]}, \ldots, Y_{2t+1,k-1}^{[g_g^o]})$. One can obtain estimates of the hyperparameters of Model–k in the same way as in **Case 1**.

The covariates, i.e., the weekday effects, are constructed by starting from "Monday*" for the $\{\mathbf{U}_t^{(k)}\}$ and "Tuesday*" for the $\{\mathbf{V}_t^{(k)}\}$, $k = 1, \ldots, p$, where the "$*$" has been added to signify that the beginning of the day has been shifted successively by $0, 1, \ldots, 23$ hours, according to which hour of day 121 is to be predicted. After removing this "weekday effect" from $\{\mathbf{U}_t^{(k)}\}$s and $\{\mathbf{V}_t^{(k)}\}$s, we use *EnviRo.stat*, free downloadable software from *http://www.enviro.stat.ubc.ca*, to implement these 24 models.

Suppose that all we have u ungauged sites in our discretized space–field field. Moreover, m out of n time points at gauged sites are formed by the unobserved response variables. Let $\mathbf{Y} = (\mathbf{Y}^{[u]}, \mathbf{Y}^{[g]}) : n \times (g+u)p$ where $\mathbf{Y}^{[g]} = (\mathbf{Y}^{[g^m]'}, \mathbf{Y}^{[g^o]'})' : n \times gp$ and $\mathbf{Y}^{[u]} : n \times up$, with $\mathbf{Y}^{[g^m]} : m \times gp$ and $\mathbf{Y}^{[g^o]} : (n-m) \times gp$. The temporal prediction problem requires the predictive posterior distribution of $(\mathbf{Y}^{[g^m]}|\mathbf{Y}^{[g^o]}, \mathcal{H})$, m being the temporal unit to be predicted. Specifically, $m = 1$ in the one–day–ahead prediction at gauged sites.

Theorem 1 *(Le and Zidek, 2006, p.160–161) Let*

$$\mathbf{Z}\beta_0^{[g]} = \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix} : \begin{pmatrix} m \times gp \\ (n-m) \times gp \end{pmatrix}$$

and

$$\mathbf{I}_n + \mathbf{Z}\mathbf{F}^{-1}\mathbf{Z}' = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} : \begin{pmatrix} m \times m & m \times (n-m) \\ (n-m) \times m & (n-m) \times (n-m) \end{pmatrix}.$$

Given all the estimated hyperparameters $\mathcal{H} = \{\mathbf{F}, \beta_0, \mathbf{\Omega}, \mathbf{\Lambda}_1, \delta_1, \mathbf{\Lambda}_0, \tau_{00}, \mathbf{H}_0, \delta_0\}$, the marginal posterior distribution is given by

$$\mathbf{Y}^{[g^m]}|\mathbf{Y}^{[g^o]}, \mathcal{H} \sim t_{m \times gp}(\mu_{(u|g)}, \mathbf{\Phi}_{(u|g)} \otimes \mathbf{\Psi}_{(u|g)}, \delta_{(u|g)}), \tag{1}$$

where

$$\mu_{(u|g)} = \mu_{(1)} + \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}(\boldsymbol{Y}^{[g^o]} - \mu_{(2)}) : m \times gp \tag{2}$$

$$\boldsymbol{\Phi}_{(u|g)} = \frac{\delta_1 - gp + 1}{\delta_1 - gp + n - m + 1}\boldsymbol{A}_{11\circ2} : m \times m \tag{3}$$

$$\boldsymbol{\Psi}_{(u|g)} = \frac{1}{\delta_1 - gp + 1}\{\boldsymbol{\Lambda}_1 \otimes \boldsymbol{\Omega} + (\boldsymbol{Y}^{[g^o]} - \mu_{(2)})'\boldsymbol{A}_{22}^{-1}(\boldsymbol{Y}^{[g^o]} - \mu_{(2)})\} : gp \times gp \tag{4}$$

$$\delta_{(u|g)} = \delta_1 - gp + n - m + 1, \tag{5}$$

where $\boldsymbol{A}_{11\circ2} = \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$.

To obtain the one–day–ahead temporal prediction at gauged sites, one needs the predictive posterior distribution of the unobserved response variable of interest, that is, the last "species" or "pollutant" in the multivariate response vector whose role is now being played by an hourly ozone concentration. Two different predictive posterior distributions of the last pollutant (i.e., the pth pollutant) are considered for Model–1 and Model–k, $k \in \{2, \ldots, p\}$. These two cases follow:

- For Model–1, $\mathbf{Y}^{[g^o]}$ has the observed responses from day 1 to day 119, and $\mathbf{Y}^{[g^m]}$ can be written as

$$\begin{aligned}
\mathbf{Y}^{[g^m]} &= ((Y^{[g_1^m]}_{121,1}, \ldots, Y^{[g_g^m]}_{121,p})', \ldots, (Y^{[g_1^o]}_{120,1}, \ldots, Y^{[g_g^o]}_{120,p})')' \\
&= \begin{pmatrix} \mathbf{Y}^{[g^m_{1:g}]}_{121,1:p} \\ \mathbf{Y}^{[g^o_{1:g}]}_{120,1:p} \end{pmatrix} : 2 \times gp,
\end{aligned}$$

with $\mathbf{Y}^{[g^m_{1:g}]}_{121,1:p} : 1 \times gp$, the unobserved response vector of day 121 and $\mathbf{Y}^{[g^o_{1:g}]}_{120,1:p} : 1 \times gp$, the observed response vector of day 120. Hence we have $m = 2$ and $n = 121$ in Theorem 1. The predictive posterior distribution of $\mathbf{Y}^{[g^m]}$ can be obtained by (2)–(5). To obtain the predictive distribution of $\mathbf{Y}^{[g^m_{1:g}]}_{121,1:p}$ given $\mathbf{Y}^{[g^o_{1:g}]}_{1:120,1:p}$, one can decompose $\mu_{(u|g)}$, $\boldsymbol{\Phi}_{(u|g)}$ and $\boldsymbol{\Psi}_{(u|g)}$ as follows:

$$\mu_{(u|g)} = \begin{pmatrix} \mu_{1r} \\ \mu_{2r} \end{pmatrix}$$

and

$$\delta_{(u|g)}\boldsymbol{\Phi}_{(u|g)} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

where $\mu_{ir} : 1 \times gp$ and $B_{ij} : 1 \times 1$ for $i, j = 1, 2$. Hence, the predictive posterior distribution of $\mathbf{Y}^{[g^m_{1:g}]}_{121,1:p}$ is given by

$$\begin{aligned}
\mathbf{Y}^{[g^m_{1:g}]}_{121,1:p} | \mathbf{Y}^{[g^o_{1:g}]}_{120,1:p}, \mathbf{Y}^{[g^o_{1:g}]}_{1:119,1:p}, \mathcal{H} \sim \; & t_{1 \times gp}(\mu_{1r} + B_{12}B_{22}^{-1}(\mathbf{Y}^{[g^o_{1:g}]}_{120,1:p} - \mu_{2r}), \frac{B_{11\circ2}}{\delta_{(u|g)} + 1} \\
& \otimes \boldsymbol{\Psi}_{(u|g)}(\mathbf{I}_{gp} + \boldsymbol{\Psi}^{-1}_{(u|g)}(\mathbf{Y}^{[g^o_{1:g}]}_{120} - \mu_{2r})'B_{22}^{-1}(\mathbf{Y}^{[g^o_{1:g}]}_{120} \\
& - \mu_{2r})), \delta_{(u|g)} + 1).
\end{aligned} \tag{6}$$

Let $\mathbf{e}_k^l : k \times 1$ denote a vector all of whose elements are 0 say for the jth which is 1, $j \neq l, j = 1, \ldots, k$. Let $\mathbf{E}_1 = $ block–diag–matrix$\{\mathbf{e}_p^p\} : gp \times g$. At Gauged Site $j \in \{1, \ldots, g\}$, the predictive distribution of the pth unobserved response $Y_{121,p}^{[g_j^m]}$, that is, $\mathbf{Y}_{121,1:p}^{[g_{1:g}^m]}\mathbf{E}_1\mathbf{e}_p^j$, also has a t–distribution:

$$Y_{121,p}^{[g_j^m]}|\mathbf{Y}_{120,1:p}^{[g_{1:g}^o]}, \mathbf{Y}_{1:119,1:p}^{[g_{1:g}^o]}, \mathcal{H} \quad \sim \quad t_{\delta_{(u|g)}+1}(\mu^*\mathbf{E}_1\mathbf{e}_p^j, \phi^*\mathbf{e}_p^{j\,\prime}\mathbf{E}_1'\mathbf{\Psi}^*\mathbf{E}_1\mathbf{e}_p^j),$$
(7)

where $\mu^* = \mu_{1r} + B_{12}B_{22}^{-1}(\mathbf{Y}_{120,1:p}^{[g_{1:g}^o]} - \mu_{2r})$, $\phi^* = \frac{B_{11\circ2}}{\delta_{(u|g)}+1}$ and $\mathbf{\Psi}^* = \mathbf{\Psi}_{(u|g)}(\mathbf{I}_{gp} + \mathbf{\Psi}_{(u|g)}^{-1}(\mathbf{Y}_{120,1:p}^{[g_{1:g}^o]} - \mu_{2r})'B_{22}^{-1}(\mathbf{Y}_{120,1:p}^{[g_{1:g}^o]} - \mu_{2r}))$.

- For Model–k, $k = 2, \ldots, p$, $\mathbf{Y}^{[g^o]}$ has the observed responses from Day 1 to Day 119, while $\mathbf{Y}^{[g^m]}$ consists of $k - 1$ unobserved responses and $p - k + 1$ observed ones at each gauged site. To predict the responses one–day–ahead at gauged sites in this field, we have $m = 1$ and $n = 120$ in Theorem 1. Let $\mathbf{E}_{2j} = (\mathbf{e}_{gp}^{(j-1)p+1}, \ldots, \mathbf{e}_{gp}^{jp}) : gp \times p$ for $j = 1, \ldots, g$. At Gauged Site (GS) $j \in \{1, \ldots, g\}$, we have

$$\mathbf{Y}^{[g^m]}\mathbf{E}_{2j}|\mathbf{Y}^{[g^o]}, \mathcal{H} \quad \sim \quad t_{1\times p}(\mu_{(u|g)}\mathbf{E}_{2j}, \Phi_{(u|g)} \otimes \mathbf{E}_{2j}'\mathbf{\Psi}_{(u|g)}\mathbf{E}_{2j}, \delta_{(u|g)}).$$
(8)

Notice that $\mathbf{Y}^{[g^m]}\mathbf{E}_{2j}$ is $(Y_{n-1,k}^{[g_j^o]}, \ldots, Y_{n-1,p}^{[g_j^o]}, Y_{n,1}^{[g_j^m]}, \ldots, Y_{n,k-1}^{[g_j^m]})$. Let $\mathbf{E}_3 = (\mathbf{e}_p^p, \ldots, \mathbf{e}_p^1) : p \times p$. Multiplying $\mathbf{Y}^{[g^m]}\mathbf{E}_{2j}$ by $\mathbf{E}_3$ reverses the order of the pollutants such that the response of the last hour being relocated in the first position of the new response vector, the response of the second last hour in the second position of the new response vector, and so on. In other words, we obtain the following new response vector: $(Y_{n,k-1}^{[g_j^m]}, \ldots, Y_{n,1}^{[g_j^m]}, Y_{n-1,p}^{[g_j^o]}, \ldots, Y_{n-1,k}^{[g_j^o]})$. That new response has the following multivariate t–distribution:

$$\mathbf{Y}^{[g^m]}\mathbf{E}_{2j}\mathbf{E}_3|\mathbf{Y}^{[g^o]}, \mathcal{H} \quad \sim \quad t_{1\times p}(\mu_j, \Phi_{(u|g)} \otimes \mathbf{\Psi}_j, \delta_{(u|g)}),$$
(9)

where $\mu_j = \mu_{(u|g)}\mathbf{E}_{2j}\mathbf{E}_3$ and $\mathbf{\Psi}_j = \mathbf{E}_3'\mathbf{E}_{2j}'\mathbf{\Psi}_{(u|g)}\mathbf{E}_{2j}\mathbf{E}_3$. Decompose $\mathbf{Y}^{[g^m]}\mathbf{E}_{2j}\mathbf{E}_3$, μ_j and $\mathbf{\Psi}_j$ as follows:

$$\mathbf{Y}^{[g^m]}\mathbf{E}_{2j}\mathbf{E}_3 = (\mathbf{T}_{1c}, \mathbf{T}_{2c}): \quad (1 \times (k-1), 1 \times (p-k+1)),$$

$$\mu_j = (\mu_{1c}, \mu_{2c}): \quad (1 \times (k-1), 1 \times (p-k+1)),$$

and

$$\mathbf{\Psi}_j = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix} : \quad \begin{pmatrix} (k-1) \times (k-1) & (k-1) \times (p-k+1) \\ (p-k+1) \times (k-1) & (p-k+1) \times (p-k+1) \end{pmatrix}.$$

Hence the unobserved response variable $\mathbf{T}_{1c}$ is also t–distributed:

$$
\begin{aligned}
\mathbf{T}_{1c}|\mathbf{T}_{2c}, \mathbf{Y}^{[g^o]}, \mathcal{H} \quad \sim \quad & t_{1\times(k-1)}(\mu_{1c} + (\mathbf{T}_{2c} - \mu_{2c})\mathbf{C}_{22}^{-1}\mathbf{C}_{21}, \frac{\delta_{(u|g)}}{\delta_{(u|g)} + p - k + 1} \\
& \times \Phi_{(u|g)}\{1 + (\delta_{(u|g)}\Phi_{(u|g)})^{-1}(\mathbf{T}_{2c} - \mu_{2c})\mathbf{C}_{22}^{-1}(\mathbf{T}_{2c} - \mu_{2c})'\} \\
& \otimes \mathbf{C}_{11\circ2}, \delta_{(u|g)} + p - k + 1).
\end{aligned}
\tag{10}
$$

The last "pollutant", that is, the first entry of $\mathbf{T}_{1c}$, can be predicted by multiplying $\mathbf{T}_{1c}$ with $\mathbf{e}_{k-1}^1$. Consequently, the predictive posterior distribution of $Y_{n,k-1}^{[g_j^m]}$ is given as follows:

$$
\begin{aligned}
Y_{n,k-1}^{[g_j^m]}|\mathbf{Y}^{[g^o]}, \mathbf{Y}_{n-1,k:p}^{[g_j^o]}, \mathcal{H} \quad \sim \quad & t_{\delta_{(u|g)}+p-k+1}((\mu_{1c} + (\mathbf{T}_{2c} - \mu_{2c})\mathbf{C}_{22}^{-1}\mathbf{C}_{21})\mathbf{e}_{k-1}^1, \\
& \frac{\delta_{(u|g)}}{\delta_{(u|g)} + p - k + 1}\Phi_{(u|g)}\{1 + (\delta_{(u|g)}\Phi_{(u|g)})^{-1} \\
& \times(\mathbf{T}_{2c} - \mu_{2c})\mathbf{C}_{22}^{-1}(\mathbf{T}_{2c} - \mu_{2c})'\}(\mathbf{e}_{k-1}^1)'\mathbf{C}_{11\circ2}\mathbf{e}_{k-1}^1).
\end{aligned}
\tag{11}
$$

The corresponding predictive variance of the $(k-1)^{\text{th}}$ hour of the 121st day at GS j is then given by:

$$
\begin{aligned}
\text{Var}(Y_{n,k-1}^{[g_j^m]}|\mathbf{Y}^{[g^o]}, \mathbf{Y}_{n-1,k:p}^{[g_j^o]}, \mathcal{H}) \quad = \quad & \frac{\delta_{(u|g)}}{\delta_{(u|g)} + p - k - 1}\Phi_{(u|g)}\{1 + (\delta_{(u|g)}\Phi_{(u|g)})^{-1} \\
& \times(\mathbf{T}_{2c} - \mu_{2c})\mathbf{C}_{22}^{-1}(\mathbf{T}_{2c} - \mu_{2c})'\}(\mathbf{e}_{k-1}^1)'\mathbf{C}_{11\circ2}\mathbf{e}_{k-1}^1.
\end{aligned}
$$

It is straightforward to construct the 95% pointwise forecast intervals at the $(k-1)^{\text{th}}$ hour of the 121st day at each gauged site from (11).

2.2 r–step–ahead prediction with the BSP

In this subsection, we generalize the forecast results with the BSP approach from one–day–ahead to r–step–ahead ($r \in \mathcal{N}$) prediction. Denote by N the total number of days of observed responses. We also consider the multivariate setting with p being the total number of pollutants or species and g, total number of gauged sites. Similarly, we generalize this forecast result for two cases: (i) predict the response variable at the last hour of the $(N+r)^{\text{th}}$ day; and (ii) predict the response variable at the $(k-1)^{\text{th}}$ hour of the $(N+r)^{\text{th}}$ day, for $k = 2, \ldots, p$.

Case (i): Predict the response variable at the last hour of the $(N+r)^{\text{th}}$ day

Lemma 1 *The odd–day–response $\{\boldsymbol{U}_t^{(1)}, t = 1, \ldots, t^O\}$ can be formed by $\boldsymbol{U}_t^{(1)} = (Y_{2t-1,1}^{[g_1^o]}, \ldots, Y_{2t-1,p}^{[g_g^o]})$. Note that $t^O = K$ if $N = 2K$ or $N = 2K-1$ for some $K \in \mathcal{N}$.*

Lemma 2 *The even–day–response* $\{\boldsymbol{V}_t^{(1)}, t = 1, \ldots, t^E\}$ *can be formed by* $\boldsymbol{V}_t^{(1)} = (Y_{2t,1}^{[g_1^o]}, \ldots, Y_{2t,p}^{[g_g^o]})$. *Note that* $t^E = K - 1$ *if* $N = 2K$ *and* $t^E = K$ *if* $N = 2K - 1$ *for some* $K \in \mathcal{N}$.

Remark 1 *Notice that the total number of observations in subdata matrices can be different, depending on whether* N *is an odd or even number. Let* t^N *be the total number of observed response variables. So* $t^N = t^O$ *for the odd–day–response and* t^E, *for the even–day–response.*

To keep things simple, we use N instead of t^N in the following theorem:

Theorem 2 *Let* $\boldsymbol{Y}^{[g^m]} = ((\boldsymbol{Y}_{N+r,1:p}^{[g_{1:g}^m]})', \ldots, (\boldsymbol{Y}_{N+1,1:p}^{[g_{1:g}^m]})', (\boldsymbol{Y}_{N,1:p}^{[g_{1:g}^o]})')' : (r + 1) \times gp$ *and* $\boldsymbol{Y}^{[g^o]} = \boldsymbol{Y}_{1:(N-1),1:p}^{[g_{1:p}^o]} : (N-1) \times gp$. *Then we have the following predictive distributions:*

(i) $(\boldsymbol{Y}^{[g^m]} | \boldsymbol{Y}^{[g^o]}, \mathcal{H}) \sim t_{(r+1) \times gp}(\breve{\mu}_{(u|g)}, \breve{\boldsymbol{\Phi}}_{(u|g)} \otimes \breve{\boldsymbol{\Psi}}_{(u|g)}, \breve{\delta}_{(u|g)})$, *where*

$$
\begin{aligned}
\breve{\mu}_{(u|g)} &= \mu_{(1)} + \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}(\boldsymbol{Y}^{[g^o]} - \mu_{(2)}) \\
\breve{\boldsymbol{\Phi}}_{(u|g)} &= \frac{\delta_1 - gp + 1}{\delta_1 - gp + N + 1}\boldsymbol{A}_{11\circ 2} \\
\breve{\boldsymbol{\Psi}}_{(u|g)} &= \frac{1}{\delta_1 - gp + 1}\{\boldsymbol{\Lambda}_1 \otimes \boldsymbol{\Omega} + (\boldsymbol{Y}^{[g^o]} - \mu_{(2)})'\boldsymbol{A}_{22}^{-1}(\boldsymbol{Y}^{[g^o]} - \mu_{(2)})\} \\
\breve{\delta}_{(u|g)} &= \delta_1 - gp + N + 1.
\end{aligned}
$$

(ii) *The predictive distribution of* $Y_{N+r,p}^{[g_j^m]}$, *the* p^{th} *unobserved response at the* $(N+r)^{th}$ *day at GS* j, *is* t–*distributed:*

$$
Y_{N+r,p}^{[g_j^m]} \sim t_{\breve{\delta}}((\boldsymbol{e}_r^j)'\breve{\mu}\boldsymbol{e}_g^j, \frac{\breve{\delta}}{\breve{\delta} - 2}(\boldsymbol{e}_r^j)'\breve{\boldsymbol{\Phi}}\boldsymbol{e}_r^j(\boldsymbol{e}_g^j)'\breve{\boldsymbol{\Psi}}\boldsymbol{e}_g^j), \tag{12}
$$

where

$$
\begin{aligned}
\breve{\mu} &= \breve{\mu}_{1r} + \breve{\boldsymbol{B}}_{12}\breve{\boldsymbol{B}}_{22}^{-1}(\boldsymbol{Y}_{N,1:p}^{[g_{1:g}^o]} - \breve{\mu}_{2r}) \\
\breve{\boldsymbol{\Phi}} &= \frac{1}{\breve{\delta}_{(u|g)} + 1}\breve{\boldsymbol{B}}_{11\circ 2} \\
\breve{\boldsymbol{\Psi}} &= \breve{\boldsymbol{\Psi}}_{(u|g)}[\boldsymbol{I}_{gp} + \breve{\boldsymbol{\Psi}}_{(u|g)}^{-1}(\boldsymbol{Y}_{N,1:p}^{[g_{1:g}^o]} - \breve{\mu}_{2r})'\breve{\boldsymbol{B}}_{22}^{-1}(\boldsymbol{Y}_{N,1:p}^{[g_{1:g}^o]} - \breve{\mu}_{2r})] \\
\breve{\delta} &= \breve{\delta}_{(u|g)} + 1.
\end{aligned}
$$

The proof and details can be seen in Appendix A.1.

Case (ii): Predict the response variable at the $(k-1)^{\text{th}}$ hour of the $(N+r)^{\text{th}}$ day, for $k = 2, \ldots, p$.

Lemma 3 *The odd–day–response $\{\boldsymbol{U}_t^{(k)}, t = 1, \ldots, t^O\}$ can be formed by $\boldsymbol{U}_t^{(k)} = (Y_{2t-1,k}^{[g_1^o]}, \ldots, Y_{2t-1,p}^{[g_1^o]}, Y_{2t,1}^{[g_1^o]}, \ldots, Y_{2t,k-1}^{[g_1^o]}, \ldots, Y_{2t-1,k}^{[g_g^o]}, \ldots, Y_{2t-1,p}^{[g_g^o]}, Y_{2t,1}^{[g_g^o]}, \ldots, Y_{2t,k-1}^{[g_g^o]}) : 1 \times gp$. Note that $t^O = K \in \mathcal{N}$ if $N = 2K$ and $t^O = K - 1$ if $N = 2K - 1$ for some $K \in \mathcal{N}$.*

Lemma 4 *The even–day–response $\{\boldsymbol{V}_t^{(k)}, t = 1, \ldots, t^E\}$ can be formed by $\boldsymbol{V}_t^{(k)} = (Y_{2t,k}^{[g_1^o]}, \ldots, Y_{2t,p}^{[g_1^o]}, Y_{2t+1,1}^{[g_1^o]}, \ldots, Y_{2t+1,k-1}^{[g_1^o]}, \ldots, Y_{2t,k}^{[g_g^o]}, \ldots, Y_{2t,p}^{[g_g^o]}, Y_{2t+1,1}^{[g_g^o]}, \ldots, Y_{2t+1,k-1}^{[g_g^o]}) : 1 \times gp$. Note that $t^E = K - 1$ if $N = 2K$ or $N = 2K - 1$ for some $K \in \mathcal{N}$.*

Remark 2 *We also let t^N represent the total number of observed responses. So t^N is t^O for an odd–day–response and t^E, for an even–day–response.*

As before we use N instead of t^N to keep notation simple.

Theorem 3 *Let $\boldsymbol{Y}^{[g^m]} = ((\boldsymbol{W}_{N+r-1}^{[g_{1:g}^m]})', \ldots, (\boldsymbol{W}_{N+1}^{[g_{1:g}^m]})', (\boldsymbol{Y}_N^{*[g_{1:g}^m]})')' : r \times gp$, where $\boldsymbol{W}_i^{[g_j^m]} = (Y_{i,k}^{[g_j^m]}, \ldots, Y_{i,p}^{[g_j^m]}, Y_{i+1,1}^{[g_j^m]}, \ldots, Y_{i+1,k-1}^{[g_j^m]}) : 1 \times gp$, and $\boldsymbol{Y}_N^{*[g_j^m]} = (Y_{N,k}^{[g_j^o]}, \ldots, Y_{N,p}^{[g_j^o]}, Y_{N+1,1}^{[g_j^m]}, \ldots, Y_{N+1,k-1}^{[g_j^m]})) : 1 \times gp$, for $i = N+1, \ldots, N+r-1$ and $j = 1, \ldots, g$. We also denote $\boldsymbol{Y}^{[g^o]} = \boldsymbol{Y}_{1:(N-1)}^{*[g_{1:g}^o]} : (N-1) \times gp$, where $\boldsymbol{Y}_i^{*[g_j^o]} = (Y_{i,k}^{[g_j^o]}, \ldots, Y_{i,p}^{[g_j^o]}, Y_{i+1,1}^{[g_j^o]}, \ldots, Y_{i+1,k-1}^{[g_j^o]}) : 1 \times gp$. We then have the following forecast distributions:*

(i)

$$\boldsymbol{Y}^{[g^m]} | Y^{[g^o]}, \mathcal{H} \quad \sim \quad t_{r \times gp}(\tilde{\mu}_{(u|g)}, \tilde{\boldsymbol{\Phi}}_{(u|g)} \otimes \tilde{\boldsymbol{\Psi}}_{(u|g)}, \tilde{\delta}_{(u|g)}),$$

where

$$\tilde{\mu}_{(u|g)} = \mu_{(1)} + \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}(\boldsymbol{Y}^{[g^o]} - \mu_{(2)}) : r \times gp$$

$$\tilde{\boldsymbol{\Phi}}_{(u|g)} = \frac{\delta_1 - gp + 1}{\delta_1 - gp + N}(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})$$

$$\tilde{\boldsymbol{\Psi}}_{(u|g)} = \frac{1}{\delta_1 - gp + 1}\{\boldsymbol{\Lambda}_1 \otimes \boldsymbol{\Omega} + (\boldsymbol{Y}^{[g^o]} - \mu_{(2)})'\boldsymbol{A}_{22}^{-1}(\boldsymbol{Y}^{[g^o]} - \mu_{(2)})\}$$

$$\tilde{\delta}_{(u|g)} = \delta_1 - gp + N.$$

(ii)

$$
(\boldsymbol{Y}^{[g^m_{1:g}]}_{(N+1):(N+r-1),1:p},\ \boldsymbol{Y}^{[g^m_{1:g}]}_{N+r,1:(k-1)}|\,\boldsymbol{Y}^{[g^o_{1:g}]}_{1:N,1:p},\mathcal{H})\ \propto\ \prod_{j=1}^{g}(\boldsymbol{T}^r_{1j}|\,\boldsymbol{Y}^{[g^m_j]}_{N+1,1:(k-1)},\ \boldsymbol{Y}^{[g^o_{1:g}]}_{1:N,1:p},\mathcal{H})
$$

$$
\times(\boldsymbol{Y}^{[g^m_j]}_{N+1,1:(k-1)}|\,\boldsymbol{Y}^{[g^o_{1:g}]}_{1:N,1:p},\mathcal{H})
$$

$$
\sim\ \prod_{j=1}^{g}t_{(r-1)\times p}(\tilde{\mu}^*_{1j},\tilde{\boldsymbol{\Phi}}^*\otimes\tilde{\boldsymbol{\Psi}}^*_j,\tilde{\delta}_{(u|g)}
$$

$$
+1)t_{1\times(k-1)}(\tilde{\mu}^*_{2j},\tilde{\boldsymbol{\Phi}}^*_{2j}\otimes\tilde{\boldsymbol{\Psi}}^*_{2j},
$$

$$
\tilde{\delta}_{(u|g)}+p-k+1),
$$

where $\tilde{\mu}^*_{1j}$, $\tilde{\boldsymbol{\Phi}}^*$, and $\tilde{\boldsymbol{\Psi}}^*_j$ are given in (15) and $\tilde{\mu}^*_{2j}$, $\tilde{\boldsymbol{\Phi}}^*_{2j}$, and $\tilde{\boldsymbol{\Psi}}^*_{2j}$, in (16).

The proof and details can be seen in Appendix A.2.

Remark 3 *From Theorem 3, the forecast distribution for the unobserved response variables from Day $N+1$ to $N+r$ is the product of a sequence of matrix–t and t distributions. This implies no analytic form can be found for the response variable at the $(k-1)^{th}$ (for $k=2,\dots,p$) hour of the $(N+r)^{th}$ day at GS j $(j=1,\dots,g)$.*

2.3 r–step–ahead prediction with the DLM

The more conventional approach of state space modelling can also be used for forecasting, more specifically in the Bayesian setting, dynamic linear modelling *DLM*. For the ozone field with which we are concerned, a spatial as well as temporal correlation structure is involved. With $\mathbf{V}$ denoting the matrix of intersite distances and λ the so–called range parameter, the measurement and state equations of the DLM can be written

$$
\begin{aligned}
\mathbf{Y}_t &= \mathbf{F}'_t\mathbf{x}_t+\nu_t & \nu_t &\sim N(\mathbf{0},\sigma^2\exp(-\mathbf{V}/\lambda))\\
\mathbf{x}_t &= \mathbf{x}_{t-1}+\omega_t & \omega_t &\sim N(\mathbf{0},\sigma^2\mathbf{W})
\end{aligned}
$$

with initial information: $\mathbf{x}_0|\mathbf{D}_0\sim N(\mathbf{m}_0,\sigma_0^2\mathbf{C}_0)$. One can obtain the posterior distribution of the state parameters at the last known time point, n, that is, $\mathbf{x}_n|\mathbf{y}_{1:n},\theta\sim N(\mathbf{m}_n,\sigma^2\mathbf{C}_n)$, using the Kalman filter, a smoothing method and the Metropolis–within–Gibbs sampling algorithm (Huerta et al. 2004; Dou et al. 2007, 2008).

Given the distribution of the state parameters at the last time point, n, the observed responses until time n, $\mathbf{y}_{1:n}$, and the model parameters, $\theta=\{\lambda,\sigma^2,a_1,a_2\}$, the r–step–ahead prediction is given by

$$
\mathbf{y}_{n+r}|\mathbf{y}_{1:n},\theta\ \sim\ N(\mathbf{F}'_{t+r}\mathbf{m}_n,\sigma^2\{\mathbf{F}'_{t+r}(\mathbf{C}_n+r\mathbf{W})\mathbf{F}_{t+r}+\exp(-\mathbf{V}/\lambda)\}),
$$

$$
(13)
$$

for $r \in \mathcal{N}$ (West and Harrison, 1998; Huerta et al., 2004). Hence, $n = 2880$ and $r = 1, \ldots, 24$ for the one–day–ahead prediction in the Chicago database. For any fixed r, the forecast response, $\mathbf{y}_{n+r}$, can also be obtained by the MCMC (Markov Chain Monte Carlo) method. More specifically, at iteration j, suppose we have updated the model parameters $\lambda^{(j)}$, $\sigma^{2(j)}$, $a_1^{(j)}$ and $a_2^{(j)}$ using the FFBS (forward–filtering–backward–sampling) algorithm (Carter and Kohn 1994; Huerta et al. 2004; Dou et al. 2007, 2008). That is, one has

$$\mathbf{x}_n | \mathbf{y}_{1:n}, \theta^{(j)} \sim N(\mathbf{m}_n^{(j)}, \sigma^{2(j)} \mathbf{C}_n^{(j)}).$$

Then, the forecast response at iteration j, $\mathbf{y}_{n+r}^{(j)}$, can be drawn from (13), that is,

$$
\begin{aligned}
\mathbf{y}_{n+r} | \mathbf{y}_{1:n}, \theta^{(j)} \sim\ & N(\mathbf{F}_{t+r}^{(j)'} \mathbf{m}_n^{(j)}, \sigma^{2(j)} \{ \mathbf{F}_{t+r}^{(j)'} (\mathbf{C}_n^{(j)} + r \mathbf{W}^{(j)}) \mathbf{F}_{t+r}^{(j)} \\
& + \exp(-\mathbf{V}/\lambda^{(j)}) \}).
\end{aligned}
$$

Consequently, the forecast responses are obtained by the sample means of $\{ \mathbf{y}_{n+r}^{(j)} : j = 1, \ldots, J \}$ ($J = 500; r = 1, \ldots, 24$). The empirical forecast intervals at the 95% nominal level can be obtained by corresponding sample quantiles.

2.4   r–step–ahead prediction with the NAIVE approach

The other alternative approach, we call NAIVE, helps us assess the model performance of the one–day–ahead prediction with the multivariate BSP and DLM approaches since it is extremely simple and hence provides a good baseline for reference. That approach models the response variable by the grand mean, day effect and hour effect. To be more specific, the response variable used in this approach is the vectorized square–root ozone levels at each gauged site. Using the same notation as above, at each gauged site $j \in \{1, \ldots, g\}$, the response variable is $\mathbf{Y}_{1:n}^j = (Y_{1,1}^{[g_j^o]}, \ldots, Y_{1,p}^{[g_j^o]}, \ldots, Y_{n,1}^{[g_j^o]}, \ldots, Y_{n,p}^{[g_j^o]})' : np \times 1$, for $n = 120$ and $p = 24$. The design matrix $\mathbf{X}$ contains three columns: the first column consists of 1s, building in long–term linear trend; the second for the days, capturing the day–of–the–week effect, i.e., Monday, Tuesday, etc.; and the last one for the hour effects, a substitute for day effect but one that gives more reasonable results. In other words, its "design matrix" can be written as

$$
\mathbf{X} = \begin{pmatrix} \mathbf{1}_p' & \mathbf{1}_p' & \epsilon_p' \\ \vdots & \vdots & \vdots \\ \mathbf{1}_p' & n\mathbf{1}_p' & \epsilon_p' \end{pmatrix} : (np) \times 3,
$$

where $\mathbf{1}_p' = (1, \ldots, 1)' : p \times 1$ and $\epsilon_p' = (1, \ldots, p)' : p \times 1$.

Then our model is given by $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$, where ε is the mean $\mathbf{0}$ Gaussian process to preserve its great simplicity. The coefficient vector at Gauged Site j, β_j, is thus estimated by the least squared estimator, $\widehat{\beta}_j = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}_{1:n}^j$.

Hence, the r–step–ahead prediction at GS j is given by $\widehat{\mathbf{Y}}^j_{n+r} = \mathbf{X}_{n+r}\widehat{\beta}_j$ for $r \in \mathcal{N}$ and $j = 1, \ldots, g$, with $\mathbf{X}_{n+r}$ being $(\mathbf{1}'_p, (n + r)\mathbf{1}'_p, \epsilon'_p) : p \times 3$.

In the next section we apply our methods to make one–day–ahead forecasts with Chicago data.

3 Application

The response of interest in this section is the square–root of hourly ground–level ozone concentrations (ppb) for an entire summer in the Chicago area, the data being extracted from the AQS database. That transformation being made to make the data distribution more nearly Gaussian. The extracted database contains 24 monitoring stations, irregularly located in this area. The hourly ozone concentrations have been measured at each of the sites considered. The joint spatial and temporal dependence of the hourly ozone levels are then modelled as a spatio–temporal process in the spatio–temporal field over the Chicago area.

To facilitate the assessment of the model's performance for interpolation and prediction, 14 sites are selected as "gauged" sites from 24 monitoring stations, the remaining 10 being taken to be ungauged sites. Figure 1 represents the geographical locations of these 14 gauged and 10 ungauged sites. Each has a few missing values but the gauged sites have many fewer zero measurements during the overall time span than most of the ungauged sites, thus providing much more information for this spatio–temporal field (see Figure 2).

Figure 3 shows a side–by–side boxplot of the square–root of hourly ozone concentrations at each one of the 24 monitoring stations across all the time points. It shows that GS 3 behaves differently because of its deviation from the median for all sites and times. Figure 1 shows that GS 3 to be near the Michigan River. However, it is unknown if the difference of the observed responses at GS 3 from the rest are due to the influence of that river or because other sites are also close to it, for example, GS 1 and 10. One might expect that any model not taking account of this difference could lead to a poor model fit. Dou et al. (2008) examined this issue when comparing the spatial interpolation of ozone concentrations' field using two different approaches: the multivariate BSP and DLM in another application. Here we omit this examination.

To explore this database further, weekday and hourly effects are examined in Figures 4 and 5, respectively. The latter are approximately constant over all gauged sites; in particular, the variability of the hourly effects from 0 A.M. to 10 A.M. is slightly larger than that of the remaining hours after 10 A.M., indicating the relatively strong constant hourly effects from 10 A.M. to 11 P.M. Also notice that the high ozone levels occur in the middle of the day. Given the starting time point of this database is 11 A.M., the peak hour of ozone levels actually occur around 4–5 P.M., consistent with the known phenomenon from other ozone studies. The weekday
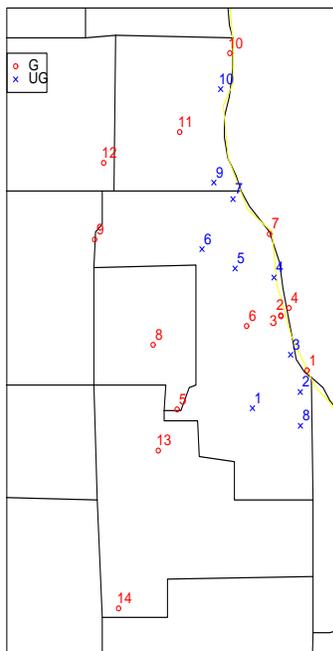
Figure 1: Geographical locations for the Chicago AQS database (2000), where the latitude and longitude are measured in degrees. ($\circ$ = G = gauged sites and $\times$ = UG = ungauged sites)

effects in Figure 4 also indicate constant weekday effects across all gauged sites. The above *exploratory data analysis* (EDA) suggests modelling constant weekday and hourly effects across all gauged sites. Constant weekday and hourly effects point to constant effects for appropriate covariates in the multivariate BSP approach. The corresponding model settings and methodology for the multivariate BSP has been discussed in the context of the Chicago area's hourly ozone concentrations' field in Section 2.1.

4 Comparisons and results

Figures 6–19 plot the forecasts of the square–root of ozone levels on the 121st day by the multivariate BSP, univariate DLM and NAIVE approaches, the 95% pointwise forecast intervals for that day by the multivariate BSP and DLM approaches, and the observations from 114th to 121st days, at each of these 14 gauged sites. The
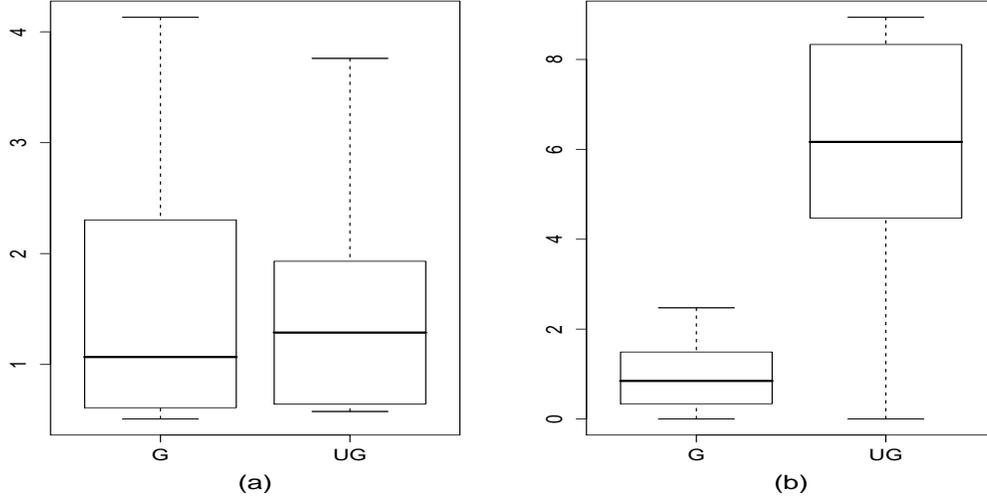
Figure 2: Boxplots for the rates of: (a) missing measurements (%); and (b) zero measurements (%), at 24 monitoring stations in the Chicago AQS database. (G = gauged sites and UG = ungauged sites)
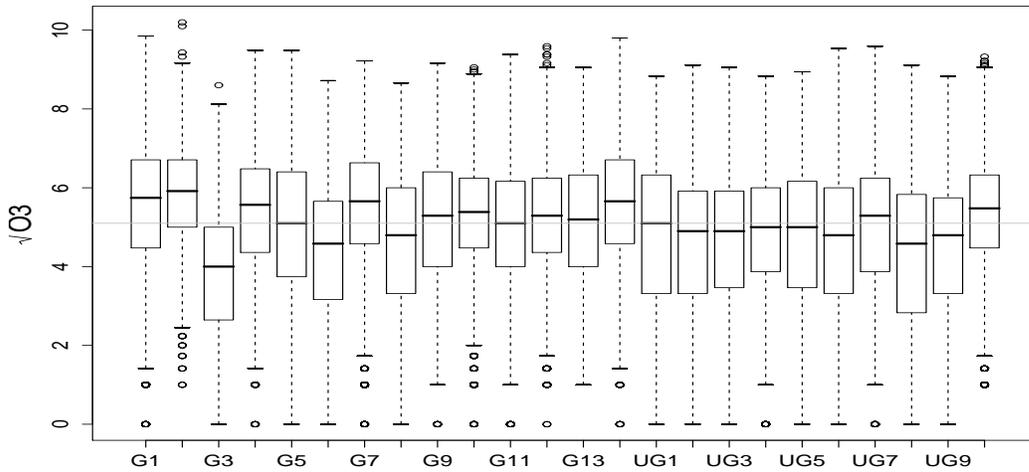
Figure 3: Boxplots for the square–root of hourly ozone concentrations ($\sqrt{ppb}$) at 24 monitoring stations in the Chicago AQS database. (G 1 = Gauged Site 1; UG 1 = Ungauged Site 1; and so on.)

multivariate BSP is much more accurate than either the DLM or NAIVE approaches. In fact, its forecast performance is rather good at most gauged sites.

Table 1 presents the mean square forecast error (MSPE) of the forecast responses

14
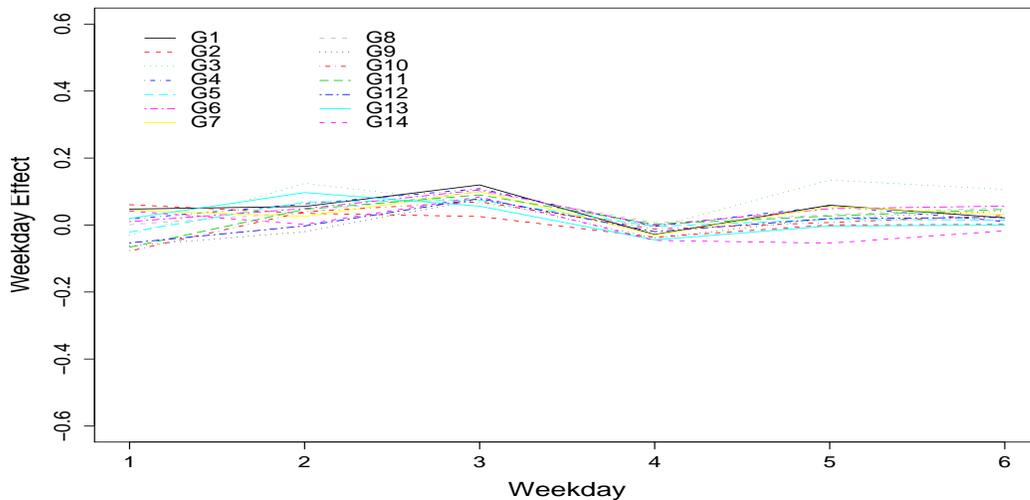
Figure 4: The weekday effect of the square–root of hourly ozone concentrations ($\sqrt{ppb}$) at 14 gauged sites in the Chicago AQS database.
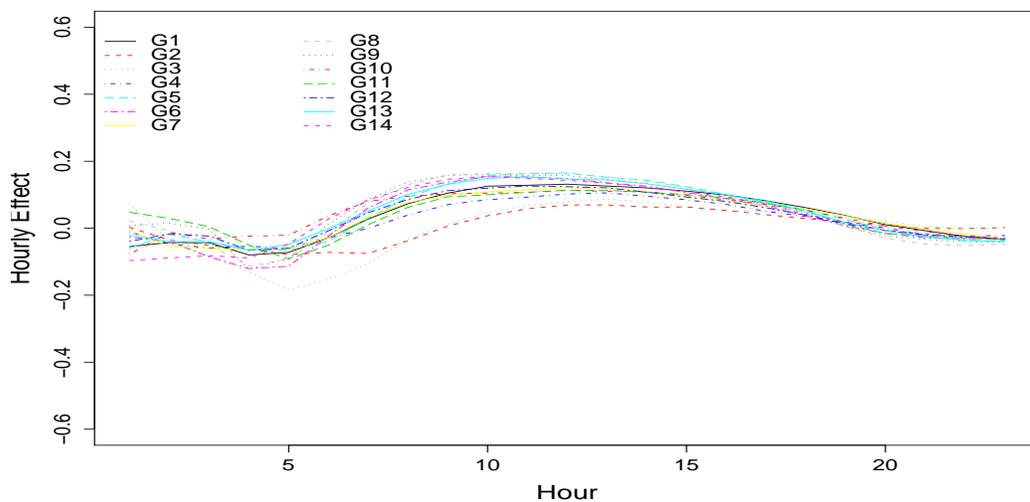
Figure 5: The hourly effect of the square–root of hourly ozone concentrations ($\sqrt{ppb}$) at 14 gauged sites in the Chicago AQS database.

on the 121st day at each one of 14 gauged sites using the three approaches. At GS j, the MSPE of the prediction at hour h can be computed by:

$$\text{MSPE}^j = \sum_{h=1}^{24} (\text{PRED}_h^j - \text{OBS}_h^j)^2,$$

15

Gauged Site	MSPE(NAIVE)	MSPE(DLM)	MSPE(BSP)
1	0.52	9.38	0.50
2	0.96	7.38	0.40
3	1.93	4.66	0.40
4	1.59	4.24	0.49
5	**2.81**	**2.60**	**3.00**
6	**0.68**	**2.31**	**0.74**
7	0.51	4.19	0.22
8	1.44	7.48	1.01
9	1.60	7.01	0.59
10	**0.44**	**5.50**	**0.50**
11	0.83	9.25	0.49
12	0.75	3.41	0.45
13	1.71	12.27	0.73
14	2.44	5.61	1.09

Table 1: The mean square forecast error (MSPE) of the one–day–ahead prediction at 14 gauged sites by the multivariate BSP, DLM, and NAIVE approaches. The BSP dominates in all but 3 cases where it essentially ties with one or another of its competitors.

where PRED_h^j is the forecast response at hour h of the 121$^{\text{st}}$ day and OBS_h^j, the observed response at the same hour of the 121$^{\text{st}}$ day, at GS j. The DLM has the poorest MSPE over all gauged sites compared with NAIVE and the BSP. The NAIVE approach performs slightly better than the BSP at GS 5, 6, and 10. The BSP carries the smallest MSPE across most gauged sites among these three.

Figure 20 plots the length of the 95% pointwise forecast intervals by the BSP at 24 hours of the 121$^{\text{st}}$ day. Starting from the middle hours of that day, i.e., 9 A.M., the forecast error bands tend to increase after that until the last hour, 11 P.M., reflecting the increasing uncertainties due to the fact that fewer responses are observed as time increases.

Figure 21 plots the length of the empirical 95% forecast intervals by the DLM at 24 hours of the 121$^{\text{st}}$ day. These lengths are close to each other but have the wiggly periodic behaviour across all gauges sites, a characteristic previously observed in Dou et al. (2007, 2008). That behaviour derives from the cosines and sines in the mean function and their random coefficients, which become components of variance when the posterior variance is found. Since these harmonic terms are squared, the 12 hour cycle has high peaks at 6 hour intervals, accounting for their wiggly nature.

Though these lengths are very close to each other, the DLM actually underestimates the forecast variabilities at gauged sites as seen in Figure 22 which plots the coverage probabilities of the DLM and BSP approaches, and also shows a slightly overestimated predictive variance for the BSP, at the 95% nominal level.

Moreover, the results in Section 2.1 have been generalized to an arbitrary time points in Section 2.2, not limited to the case of 121 days of response vectors in this paper.

Therefore, we conclude that the multivariate BSP approach is more accurate on the one–day–ahead prediction at gauged sites in the Chicago AQS database (2000) than both the NAIVE and DLM approaches.
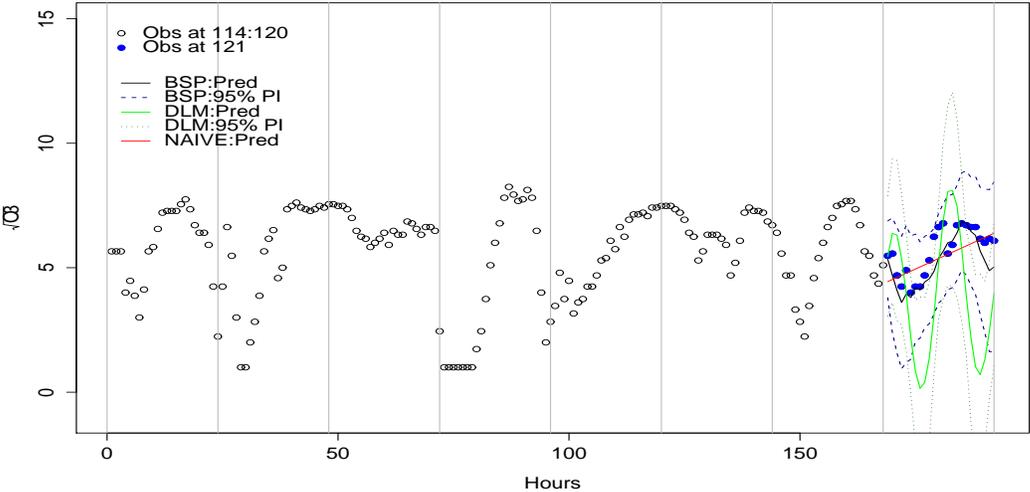
Figure 6: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from Day 114 to Day 121, the predicted values using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise forecast intervals using the multivariate BSP and DLM approaches at GS 1.
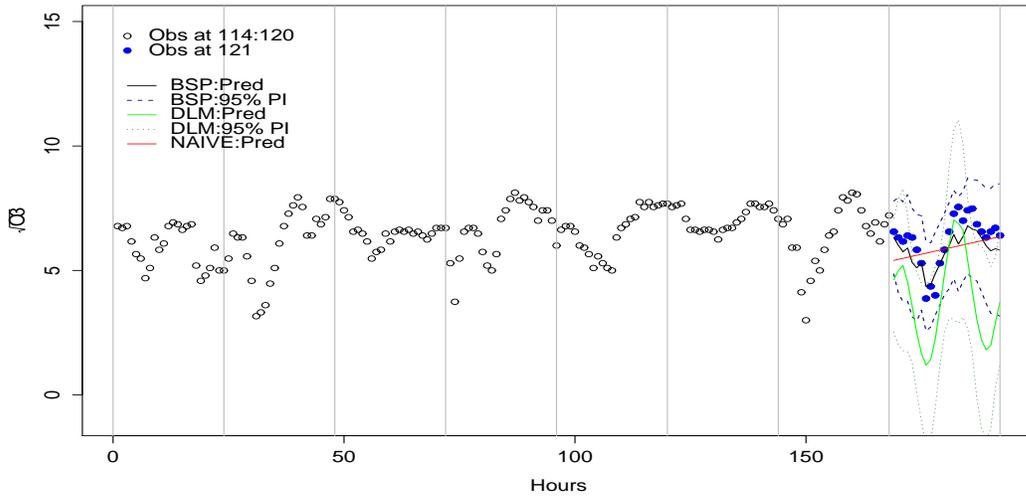
Figure 7: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from Day 114 to Day 121, the predicted values using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise forecast intervals using the multivariate BSP and DLM approaches at GS 2.
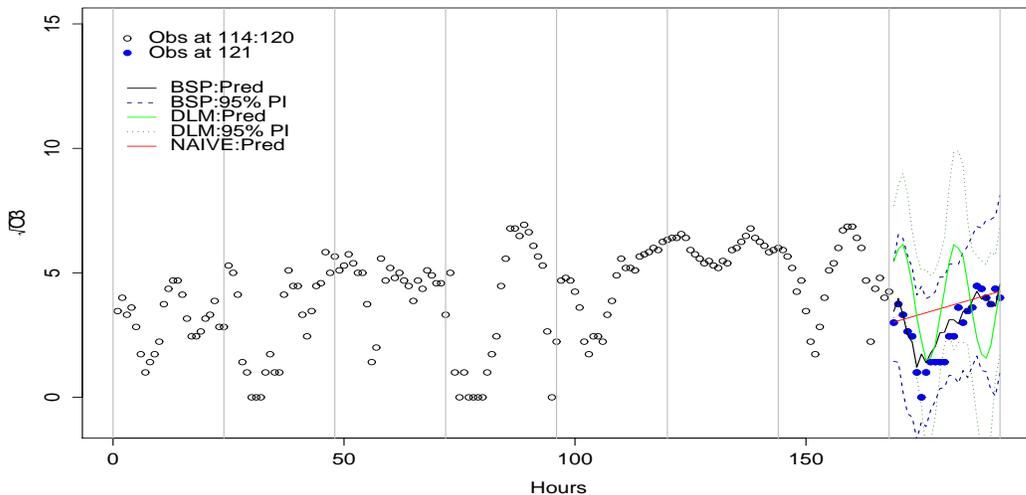
Figure 8: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from Day 114 to Day 121, the predicted values using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise forecast intervals using the multivariate BSP and DLM approaches at GS 3.
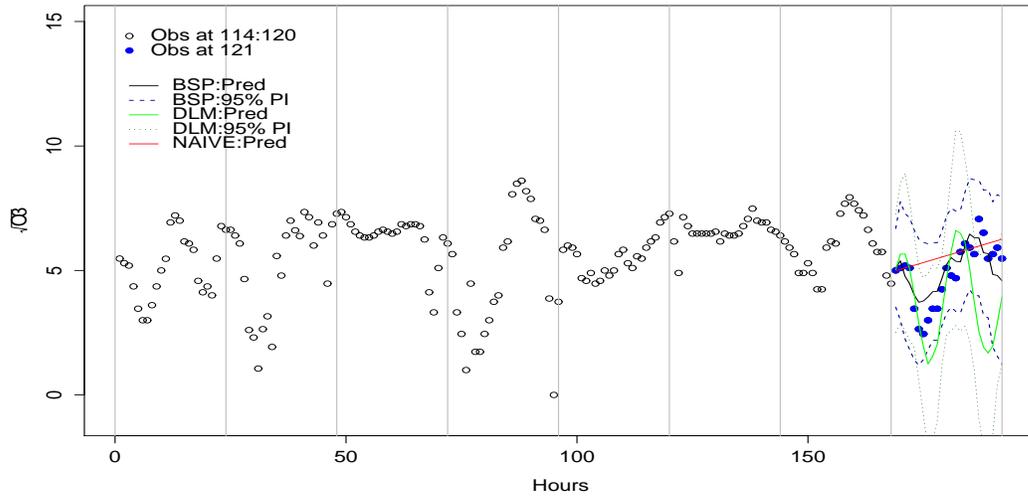
Figure 9: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from Day 114 to Day 121, the predicted values using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise forecast intervals using the multivariate BSP and DLM approaches at GS 4.
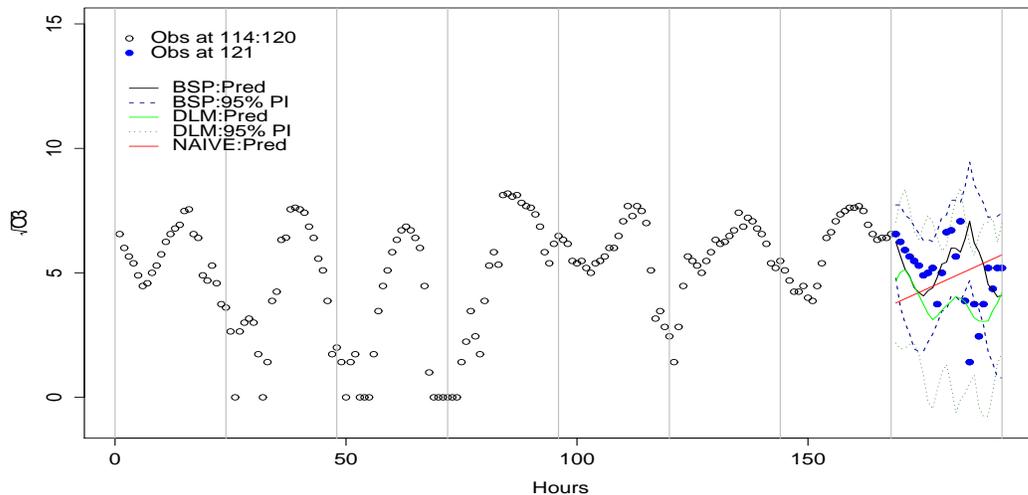
Figure 10: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from Day 114 to Day 121, the predicted values using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise forecast intervals using the multivariate BSP and DLM approaches at GS 5.
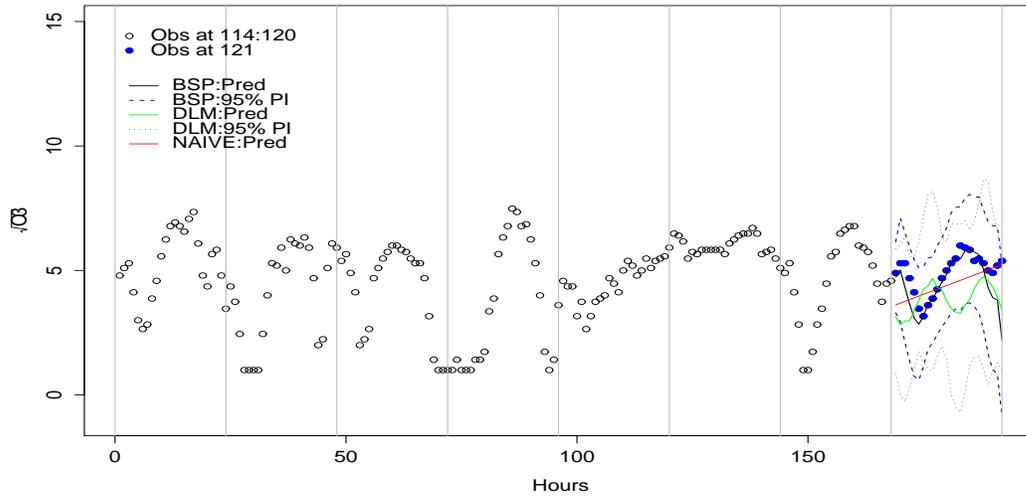
Figure 11: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from Day 114 to Day 121, the predicted values using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise forecast intervals using the multivariate BSP and DLM approaches at GS 6.
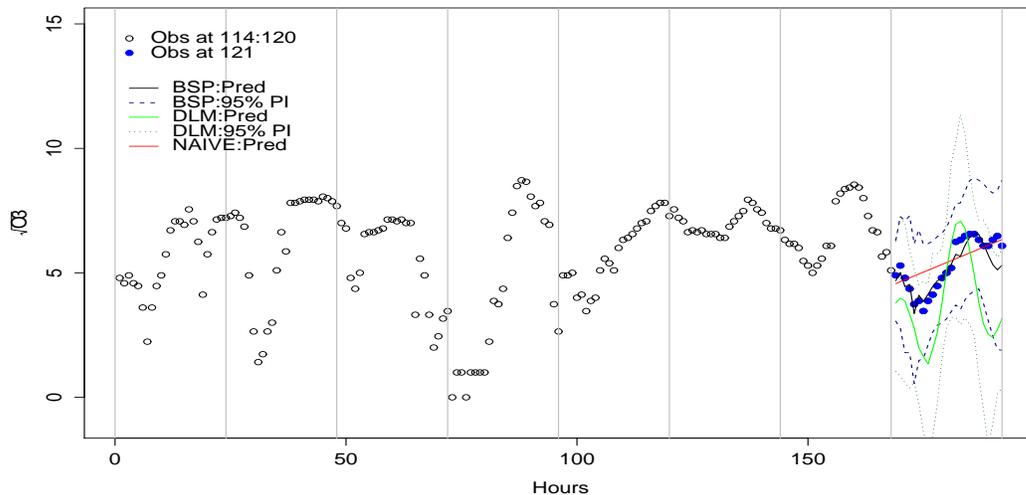
Figure 12: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from Day 114 to Day 121, the predicted values using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise forecast intervals using the multivariate BSP and DLM approaches at GS 7.
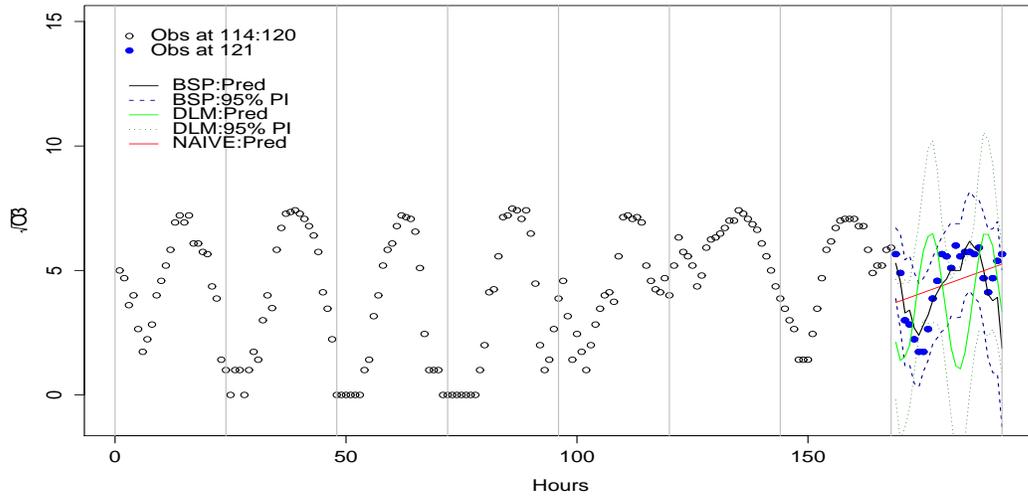
Figure 13: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from Day 114 to Day 121, the predicted values using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise forecast intervals using the multivariate BSP and DLM approaches at GS 8.
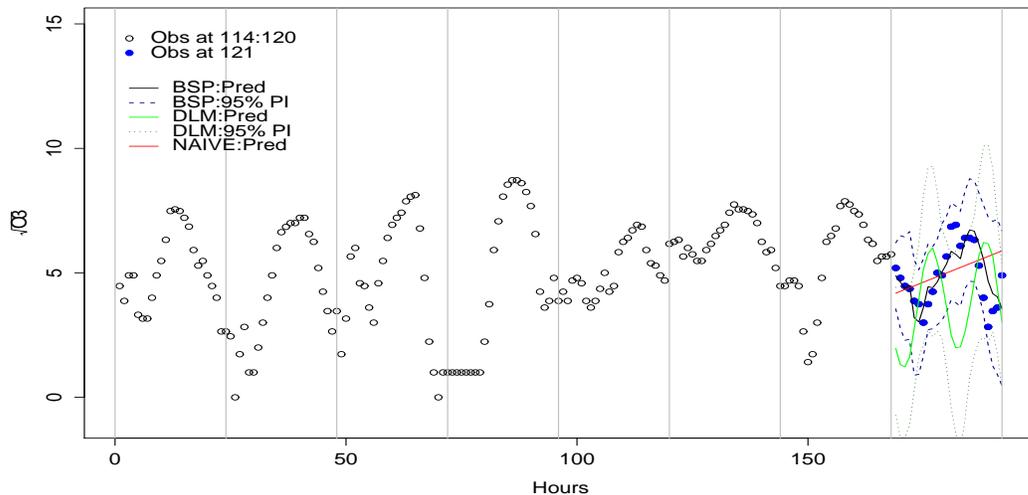
Figure 14: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from Day 114 to Day 121, the predicted values using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise forecast intervals using the multivariate BSP and DLM approaches at GS 9.
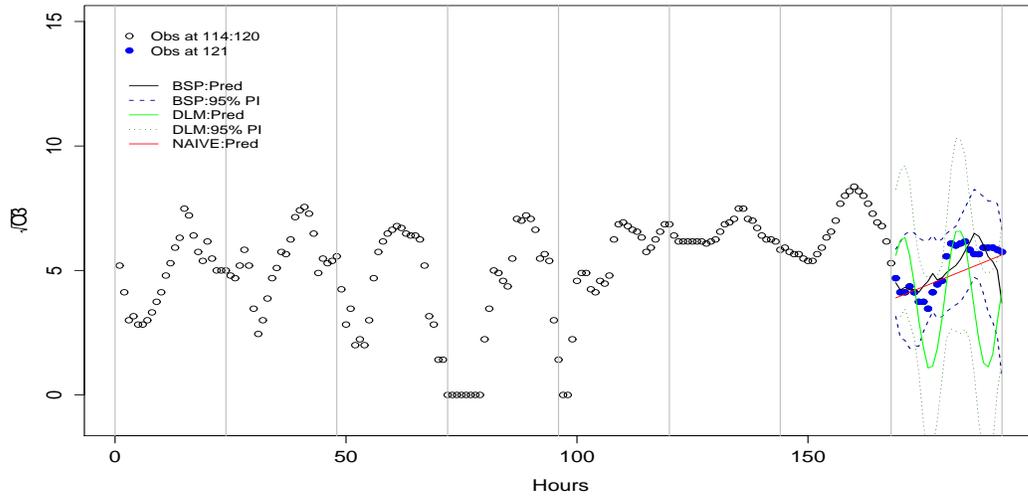
Figure 15: The observed square–root of ozone concentrations ($\sqrt{\mathrm{ppb}}$) from Day 114 to Day 121, the predicted values using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise forecast intervals using the multivariate BSP and DLM approaches at GS 10.
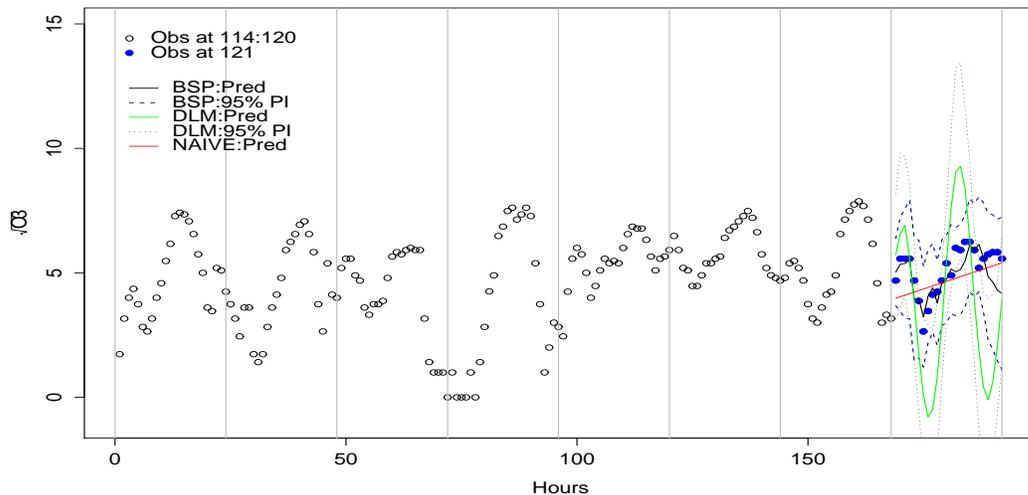
Figure 16: The observed square–root of ozone concentrations ($\sqrt{\mathrm{ppb}}$) from Day 114 to Day 121, the predicted values using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise forecast intervals using the multivariate BSP and DLM approaches at GS 11.
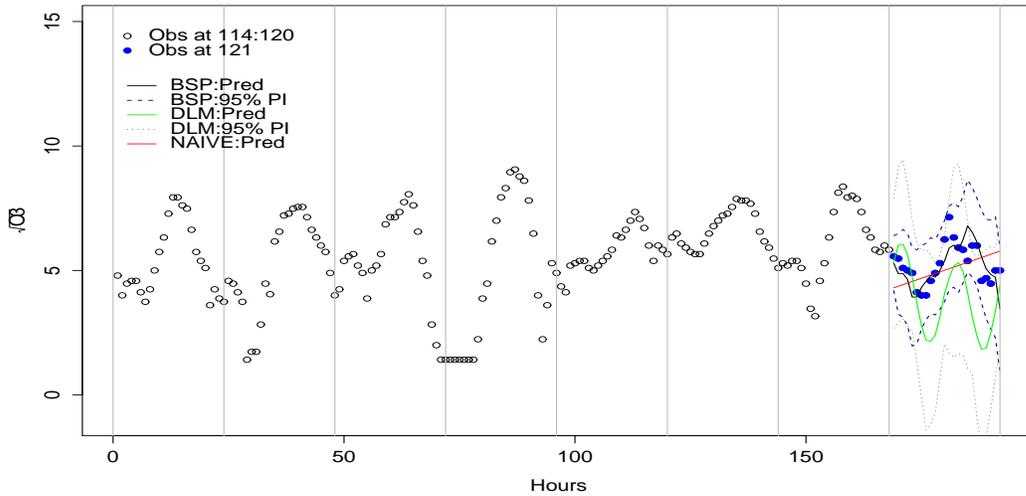
Figure 17: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from Day 114 to Day 121, the predicted values using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise forecast intervals using the multivariate BSP and DLM approaches at GS 12.
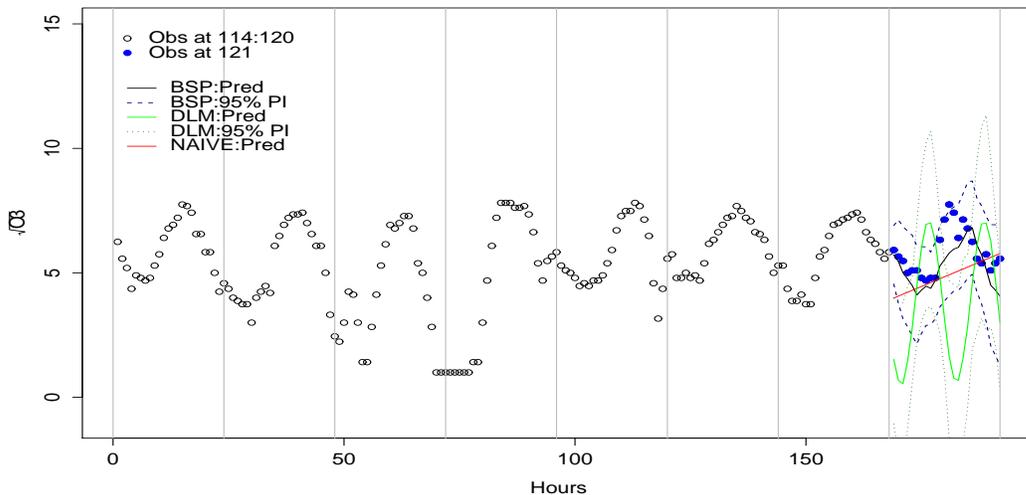
Figure 18: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from Day 114 to Day 121, the predicted values using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise forecast intervals using the multivariate BSP and DLM approaches at GS 13.
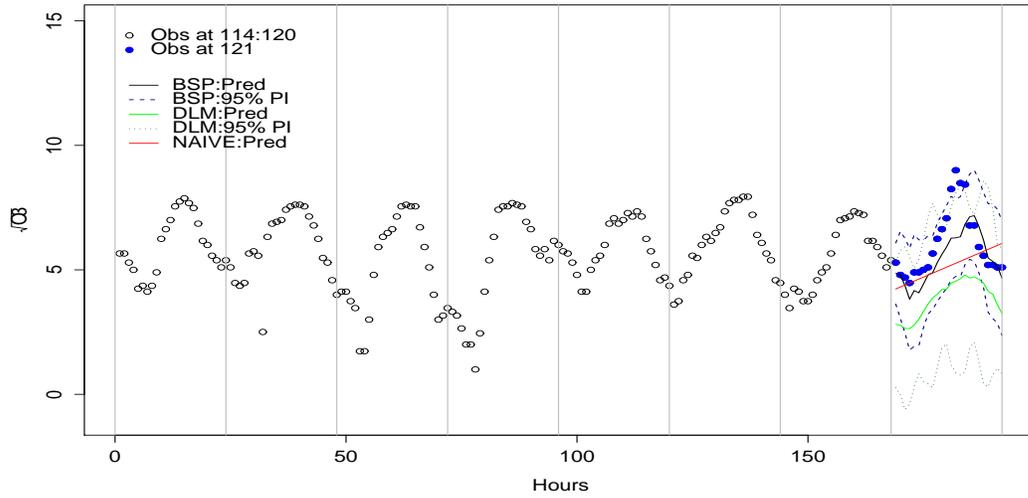
Figure 19: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from Day 114 to Day 121, the predicted values using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise forecast intervals using the multivariate BSP and DLM approaches at GS 14.
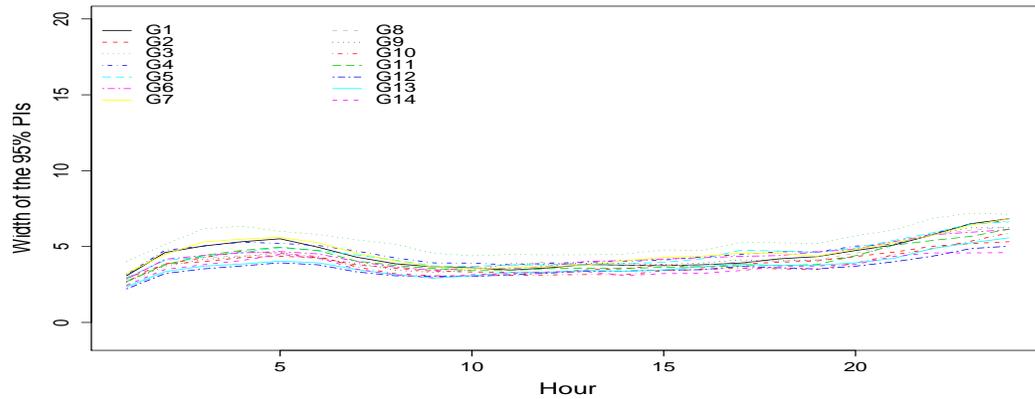
Figure 20: The width of the 95% pointwise forecast intervals of the one–day–ahead prediction at 14 gauged sites using the multivariate BSP approach.
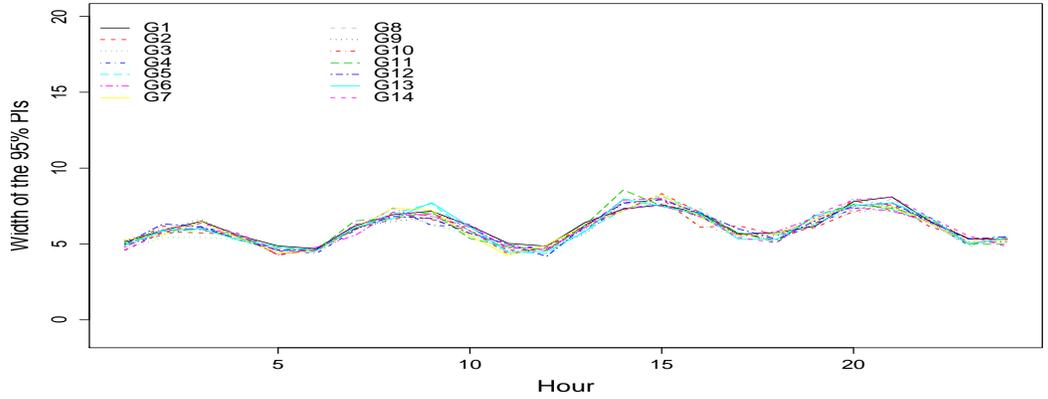
Figure 21: The width of the 95% pointwise forecast intervals (PIs) of the one–day–ahead prediction at 14 gauged sites using the DLM approach.
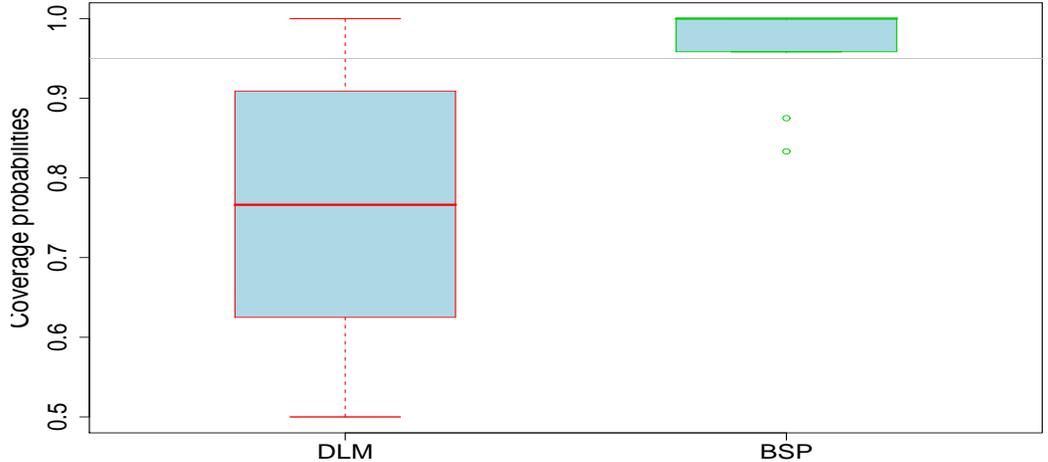
Figure 22: Boxplots of the coverage probabilities using the DLM and multivariate BSP approaches at the 95% nominal level.

5 Conclusion

The temporal prediction of the ground–level ozone concentrations in the Chicago's hourly ozone field shows the success of the adjusted multivariate BSP approach, comparing with two others: the DLM and NAIVE. This approach provides a promising new approach, currently under investigation, to time series analysis in complicated

25

situations like the one addressed in this paper where the small time scale correlations vary over time and are hard to model by conventional approaches.

A potential problem with this adaptive approach is due to the loss of the information when only a subset of the whole database is used. Further extension of the correlated response vector in the multivariate BSP modelling needs to be explored. One possible solution is a dynamic version of the multivariate BSP, the topic for a manuscript currently under preparation (Dou et al 2009).

A   Supplementary results

A.1   Results for Theorem 2

Proof 1   *(i) The result is straightforward by Theorem 1 where $m = r + 1$ and $n = N + r$.*

(ii) Decompose $\breve{\mu}_{(u|g)}$ and $\delta_{(u|g)}\breve{\Phi}_{(u|g)}$ as follows:

$$\breve{\mu}_{(u|g)} = \left(\begin{array}{c} \breve{\mu}_{1r} \\ \breve{\mu}_{2r} \end{array} \right) : \left(\begin{array}{c} r \times gp \\ 1 \times gp \end{array} \right)$$

and

$$\delta_{(u|g)}\breve{\Phi}_{(u|g)} = \left(\begin{array}{cc} \breve{B}_{11} & \breve{B}_{12} \\ \breve{B}_{21} & \breve{B}_{22} \end{array} \right) : \left(\begin{array}{cc} r \times r & r \times 1 \\ 1 \times r & 1 \times 1 \end{array} \right).$$

Hence, we have

$$\boldsymbol{Y}^{[g_{1:g}^m]}_{(N+1):(N+r),1:p} \big| \, \boldsymbol{Y}^{[g_{1:g}^o]}_{N,1:p}, \, \boldsymbol{Y}^{[g_{1:g}^o]}_{1:(N-1),1:p}, \mathcal{H} \quad \sim \quad t_{r \times gp}(\breve{\mu}, \breve{\Phi} \otimes \breve{\Psi}, \breve{\delta}), \qquad (14)$$

where $\breve{\mu}$, $\breve{\Phi}$, $\breve{\Psi}$ and $\breve{\delta}$ are given in Theorem 2.

We have $(\boldsymbol{e}_r^j)' \boldsymbol{Y}^{[g_{1:g}^m]}_{(N+1):(N+r),1:p} \boldsymbol{E}_1 \boldsymbol{e}_g^j = Y^{[g_j^m]}_{N+r,p}$, that is, the unobserved response of the last hour of the $(N+r)^{th}$ day at Gauged Site j $(j = 1, \ldots, g)$. Hence we have

$$Y^{[g_j^m]}_{N+r,p} \quad \sim \quad t_{1 \times 1}((\boldsymbol{e}_r^j)' \breve{\mu} \boldsymbol{e}_g^j, (\boldsymbol{e}_r^j)' \breve{\Phi} \boldsymbol{e}_r^j \otimes ((\boldsymbol{e}_g^j)' \breve{\Psi} \boldsymbol{e}_g^j), \breve{\delta}),$$

that is, $t_{\breve{\delta}}((\boldsymbol{e}_r^j)' \breve{\mu} \boldsymbol{e}_g^j, \frac{\breve{\delta}}{\breve{\delta}-2}(\boldsymbol{e}_r^j)' \breve{\Phi} \boldsymbol{e}_r^j (\boldsymbol{e}_g^j)' \breve{\Psi} \boldsymbol{e}_g^j)$.

A.2   Results for Theorem 3

Proof 2   *(i) The result is straightforward by Theorem 1 where $m = r$ and $n = N + r - 1$.*

26

(ii) Same as in Section 2.1, we have

$$\tilde{\boldsymbol{Y}}^{[g_j^m]} = \boldsymbol{Y}^{[g^m]}\boldsymbol{E}_{2j}\boldsymbol{E}_3 = \begin{pmatrix} Y^{[g_j^m]}_{N+r,k-1} & \cdots & Y^{[g_j^m]}_{N+r,1} & Y^{[g_j^m]}_{N+r-1,1} & \cdots & Y^{[g_j^m]}_{N+r-1,k} \\ \vdots & & \vdots & \vdots & & \vdots \\ Y^{[g_j^m]}_{N+1,k-1} & \cdots & Y^{[g_j^m]}_{N+1,1} & Y^{[g_j^o]}_{N,1} & \cdots & Y^{[g_j^o]}_{N,k} \end{pmatrix} : r \times p,$$

for $j = 1, \ldots, g$. From (i) in Theorem 3, we have

$$\tilde{\boldsymbol{Y}}^{[g_j^m]} \sim t_{r \times p}(\tilde{\mu}_j, \tilde{\boldsymbol{\Phi}}_{(u|g)} \otimes \tilde{\boldsymbol{\Psi}}_j, \tilde{\delta}_{(u|g)}),$$

where

$$\begin{aligned} \tilde{\mu}_j &= \tilde{\mu}_{(u|g)}\boldsymbol{E}_{2j}\boldsymbol{E}_3 \\ \tilde{\boldsymbol{\Psi}}_j &= \boldsymbol{E}_3'\boldsymbol{E}_{2j}'\tilde{\boldsymbol{\Psi}}_{(u|g)}\boldsymbol{E}_{2j}\boldsymbol{E}_3. \end{aligned}$$

We first decompose $\tilde{\boldsymbol{Y}}^{[g_j^m]}$, $\tilde{\mu}_j$ and $\tilde{\delta}_{(u|g)}\tilde{\boldsymbol{\Phi}}_{(u|g)}$ as follows:

$$\begin{aligned} \tilde{\boldsymbol{Y}}^{[g_j^m]} &= \begin{pmatrix} \boldsymbol{T}_{1j}^r \\ \boldsymbol{T}_{2j}^r \end{pmatrix} &:& \begin{pmatrix} (r-1) \times p \\ 1 \times p \end{pmatrix} \\ \tilde{\mu}_j &= \begin{pmatrix} \tilde{\mu}_{1j} \\ \tilde{\mu}_{2j} \end{pmatrix} &:& \begin{pmatrix} (r-1) \times p \\ 1 \times p \end{pmatrix} \\ \tilde{\delta}_{(u|g)}\tilde{\boldsymbol{\Phi}}_{(u|g)} &= \begin{pmatrix} \tilde{\boldsymbol{\Phi}}_{11} & \tilde{\boldsymbol{\Phi}}_{12} \\ \tilde{\boldsymbol{\Phi}}_{21} & \tilde{\boldsymbol{\Phi}}_{22} \end{pmatrix} &:& \begin{pmatrix} (r-1) \times (r-1) & (r-1) \times 1 \\ 1 \times (r-1) & 1 \times 1 \end{pmatrix}. \end{aligned}$$

Consequently, we have

(a)
$$\boldsymbol{T}_{2j}^r \mid \boldsymbol{Y}^{[g^o]}, \mathcal{H} \sim t_{1 \times p}(\tilde{\mu}_{2j}, \tilde{\boldsymbol{\Phi}}_{22} \otimes \tilde{\boldsymbol{\Psi}}_j, \tilde{\delta}_{(u|g)})$$

(b)
$$\boldsymbol{T}_{1j}^r \mid \boldsymbol{T}_{2j}^r, \boldsymbol{Y}^{[g^o]}, \mathcal{H} \sim t_{(r-1) \times p}(\tilde{\mu}_{1j}^*, \tilde{\boldsymbol{\Phi}}^* \otimes \tilde{\boldsymbol{\Psi}}_j^*, \tilde{\delta}_{(u|g)} + 1),$$

   where

$$\begin{cases} \tilde{\mu}_{1j}^* &= \tilde{\mu}_{1j} + \tilde{\boldsymbol{\Phi}}_{12}\tilde{\boldsymbol{\Phi}}_{22}^{-1}(\boldsymbol{T}_{2j}^r - \tilde{\mu}_2) \\ \tilde{\boldsymbol{\Phi}}^* &= \frac{\tilde{\delta}_{(u|g)}}{\tilde{\delta}_{(u|g)}+1}(\tilde{\boldsymbol{\Phi}}_{11} - \tilde{\boldsymbol{\Phi}}_{12}\tilde{\boldsymbol{\Phi}}_{22}^{-1}\tilde{\boldsymbol{\Phi}}_{21}) \\ \tilde{\boldsymbol{\Psi}}_j^* &= \tilde{\boldsymbol{\Psi}}_j(\boldsymbol{I}_p + \tilde{\boldsymbol{\Psi}}_j^{-1}(\boldsymbol{T}_{2j}^r - \tilde{\mu}_{2j})'\tilde{\boldsymbol{\Phi}}_{22}^{-1}(\boldsymbol{T}_{2j}^r - \tilde{\mu}_{2j})). \end{cases} \tag{15}$$

We then decompose $\boldsymbol{T}_{2j}^r$, $\tilde{\mu}_{2j}^r$, and $\tilde{\boldsymbol{\Psi}}_j$ as follows:

$$\begin{aligned} \boldsymbol{T}_{2j}^r &= (\boldsymbol{T}_{21}^j \quad \boldsymbol{T}_{22}^j) &:& (1 \times (k-1) \quad 1 \times (p-k+1)) \\ \tilde{\mu}_{2j}^r &= (\mu_{21}^j \quad \mu_{22}^j) &:& (1 \times (k-1) \quad 1 \times (p-k+1)) \\ \tilde{\boldsymbol{\Psi}}_j &= \begin{pmatrix} \tilde{\boldsymbol{\Psi}}_{11}^j & \tilde{\boldsymbol{\Psi}}_{12}^j \\ \tilde{\boldsymbol{\Psi}}_{21}^j & \tilde{\boldsymbol{\Psi}}_{22}^j \end{pmatrix} &:& \begin{pmatrix} (k-1) \times (k-1) & (k-1) \times (p-k+1) \\ (p-k+1) \times (k-1) & (p-k+1) \times (p-k+1) \end{pmatrix}. \end{aligned}$$

27

Hence the predictive distribution of $\boldsymbol{T}_{21}^j$, i.e., $\boldsymbol{Y}_{N+1,1:(k-1)}^{[g_j^m]}$ is given by

$$\boldsymbol{Y}_{N+1,1:(k-1)}^{[g_j^m]} \mid \boldsymbol{Y}^{[g^o]}, \mathcal{H} \sim t_{1\times(k-1)}\big(\tilde{\mu}_{2j}^*, \tilde{\boldsymbol{\Phi}}_{2j}^* \otimes \tilde{\boldsymbol{\Psi}}_{2j}^*, \tilde{\delta}_{(u|g)} + p - k + 1\big),$$

where

$$\begin{cases} \tilde{\mu}_{2j}^* &= \mu_{21}^j + (\boldsymbol{T}_{22}^j - \mu_{22}^j)(\tilde{\boldsymbol{\Psi}}_{22}^j)^{-1}\tilde{\boldsymbol{\Psi}}_{21}^j \\ \tilde{\boldsymbol{\Phi}}_{2j}^* &= \frac{\tilde{\delta}_{(u|g)}}{\tilde{\delta}_{(u|g)}+p-k+1}\tilde{\boldsymbol{\Phi}}_{22}^j\big(1 + (\tilde{\delta}_{(u|g)}\tilde{\boldsymbol{\Phi}}_{22}^j)^{-1}(\boldsymbol{T}_{22}^j - \mu_{22}^j)(\tilde{\boldsymbol{\Psi}}_{22}^j)^{-1}(\boldsymbol{T}_{22}^j - \mu_{22}^j)'\big) \\ \tilde{\boldsymbol{\Psi}}_2^* &= \tilde{\boldsymbol{\Psi}}_{11}^j - \tilde{\boldsymbol{\Psi}}_{12}^j(\tilde{\boldsymbol{\Psi}}_{22}^j)^{-1}\tilde{\boldsymbol{\Psi}}_{21}^j. \end{cases}$$

$$(16)$$

Therefore, we have

$$\begin{aligned} p\big(\boldsymbol{Y}_{(N+1):(N+r-1),1:p}^{[g_{1:g}^m]}, \boldsymbol{Y}_{N+r,1:(k-1)}^{[g_{1:g}^m]} \mid \boldsymbol{Y}_{1:N,1:p}^{[g_{1:g}^o]}, \mathcal{H}\big) \;\; &\propto \;\; \prod_{j=1}^{g} p\big(\boldsymbol{T}_{1j}^r \mid \boldsymbol{Y}_{N+1,1:(k-1)}^{[g_j^m]}, \boldsymbol{Y}_{1:N,1:p}^{[g_{1:g}^o]}, \mathcal{H}\big) \\ &\qquad \times p\big(\boldsymbol{Y}_{N+1,1:(k-1)}^{[g_j^m]} \mid \boldsymbol{Y}_{1:N,1:p}^{[g_{1:g}^o]}, \mathcal{H}\big). \end{aligned}$$

The predictive distribution for $\boldsymbol{Y}_{N+r,k-1}^{[g_{1:g}^m]}$ has no analytic form.

References

[1] Brown, PJ, Le, ND and Zidek, JV (1994). Multivariate spatial interpolation and exposure to air pollutants, *Can. J. Statist.*, 22, 489–510.

[2] Carter, CK and Kohn R (1994). On gibbs sampling for state space models. *Biometrika*, 81, 541–553.

[3] Dou, YP, Le, ND and Zidek, JV (2007). A dynamic linear model for hourly ozone concentrations. *Technical Report #228*, Department of Statistics, UBC.

[4] Dou, YP, Le, ND and Zidek, JV (2008). Modeling hourly ozone concentration fields. *Submitted.*

[5] Dou, YP, Le, ND and Zidek, JV (2009). Generalized Bayesian spatial methodology. *In preparation.*

[6] Huerta, G, Sanso, B, and Stroud, JR (2004). A spatio–temporal model for Mexico city ozone levels. *J. R. Statist. Soc. C*, 53, 231–248.

[7] Le, ND, Sun, W and Zidek, JV (1997). Bayesian multivariate spatial interpolation with data missing by design. *J. Roy. Stat. Soc.*, Ser B, 59, 501–510.

[8] Le, ND and Zidek, JV (1992). Interpolation with uncertain spatial covariance: a Bayesian alternative to kriging. *J. Mult. Anal.*, 43, 351–374.

[9] Le, N.D., & Zidek, J.V. (2006). *Statistical analysis of environmental space–time processes.* Springer.

[10] Zidek, JV, Sun, L, Le, ND, and Özkaynak, H (2002). Contending with space–time interaction in the spatial prediction of pollution: Vancouver's hourly ambient PM_{10} field. *Environmetrics*, 13, 1–19.