DYNAMIC BAYESIAN MODELS FOR MODELLING ENVIRONMENTAL SPACE–TIME FIELDS

by

Yiping Dou

B.Sc., The XuZhou Normal University, 1997
M.Sc., The University of New Brunswick, 2003
Ph.D., The University of British Columbia, 2008

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

The Faculty of Graduate Studies

(Statistics)

The University Of British Columbia

(Vancouver)

March, 2008

Abstract

This thesis addresses spatial interpolation and temporal prediction using air pollution data by several space–time modelling approaches. Firstly, we implement the dynamic linear modelling (DLM) approach in spatial interpolation and find various potential problems with that approach. We develop software to implement our approach. Secondly, we implement a Bayesian spatial prediction (BSP) approach to model spatio–temporal ground–level ozone fields and compare the accuracy of that approach with that of the DLM. Thirdly, we develop a Bayesian version empirical orthogonal function (EOF) method to incorporate the uncertainties due to temporally varying spatial process, and the spatial variations at broad– and fine– scale. Finally, we extend the BSP into the DLM framework to develop a unified Bayesian spatio–temporal model for univariate and multivariate responses. The result generalizes a number of current approaches in this field.

Table of Contents

Appendices

List of Tables

List of Figures

Acknowledgements

I am most grateful to my co–supervisors, Constance van Eeden, Nhu Le and Jim Zidek, the financial support they provided as well as the flexible working environment they provided. I greatly appreciated the encouragement provided by all three of these individuals.

I benefited from my numerous discussions with Jim, who provided inspiration and encouraged my growth through his excellence guidance throughout my PhD studies. Jim's patience, valuable suggestions and comments contributed greatly to making this thesis feasible. I am honoured to have the chance to work with him.

Discussions with Nhu Le proved invaluable and I benefited greatly from his comments and his inspired suggestions. In particular, I profited from his very perceptive insights on how my work related to the realities I set out to describe.

I cannot thank Constance enough for her wise counsel, especially on her advice on technical writing and teaching. She constantly encouraged me both academically and personally.

I would like to thank the other member of my advisory committee, Paul Gustafson, for his insightful comments and suggestions. I would like to extend my thanks to all my thesis committee for their prompt responses and suggestions on my thesis during the holiday.

I am indebted for Harry Joe and Weiliang Qiu for their help on providing some of the C codes used in the software that has been developed in this thesis. I would like to thank Nancy Heckman for her invaluable suggestions when I was trying to solve a problem in my research, and Rick White for his advice on using the C++ language.

I would like to thank Prasad Kasibhatla of Nicholas School of Environ-

To Xiulan, Benju, Huazhu, and Yuanlie

Chapter 1

Introduction

This thesis addresses one topic in four themes. More specifically, we develop a unified fully Bayesian hierarchical modelling approach to the interpolation and prediction of univariate (respectively multivariate) response variables (respectively vectors) in spatial–temporal fields, or space–time fields. The importance of the prediction of certain responses turn out to be vital for human health. Moreover, people tend to find faster, cheaper and better method for prediction. These all motivates studies in this thesis.

Many researchers in these areas face responses with spatial structure that changes over time in a dynamic fashion in the sense that the underlying process varies in space and time (Wikle and Royle, 2004). The associated random field is called a "space–time field". Measurements taken on that field yield so–called spatio–temporal data because of stochastic dependence relationships that are both spatial and temporal in nature. Cressie (1993) defines a space–time field as a set of stochastic processes over space that differ over time. Space–time modelling requires that we deal with space–time data for a variety of purposes, using various methods.

The ubiquity of such processes has led to a rich research literature on space–time data problems spread over diverse fields such as environmental health, climatology, epidemiology and ecology. For example, we may wish to investigate the relationship between air pollution and health outcomes, such as, asthma or chronic obstructive pulmonary disease. We may record the measurements at a number of monitoring sites within the study region. Each of those sites may: measure different sets of pollutants; contain missing data; have a startup time that differs from those of other monitoring sites in the network.

To model such data, we use a stochastic space–time model to capture the

dependence between pollutants, spatially and temporally. That dependence may derive from other processes than just those associated with the pollutants themselves. For example, the dependence between pollutants could be due to wind direction and speed since some pollutants such as the ground–level ozone concentrations may spread with the wind. Temporal dependence may be due to the atmospheric processes that generate the pollutants. Spatial dependence can yield a high correlation between the responses for each pollutant at different monitoring sites. Monitors in geographical proximity tend to be highly correlated unlike those that are far apart. These correlations among the monitoring sites in our application seem relatively constant over time.

Section 1.1 introduces the AQS database used in this thesis and some features of ground–level ozone concentrations. Section 1.2 describes some background literature on the topic this thesis addresses. Section 1.3 presents the plan of this thesis.

1.1   AQS Database and Ground–level Ozone

The Air Quality System (AQS) database contains measurements of air pollutant concentrations in the United States, for both *criteria air pollutants* and *hazardous air pollutants*. The former are of more concern as they are regulated under the US Clean Air Act of 1970 to protect human health and welfare. The US Environmental Protection Agency (EPA) set air quality standards for six criteria air pollutants: Carbon Monoxide (CO), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), Ozone (O3), Particulate Matter (PM10 and PM2.5), and Lead (Pb).

Ozone is *good up bad down* in terms of the environment and health, that is, ozone could be good or bad depending on its location in the atmosphere. "Good" ozone occurs in the high altitude, about 6 to 30 miles from the surface of the Earth, also called *stratosphere ozone*. Stratosphere ozone helps reduce the harmful ultraviolet (UV) rays to protect life on Earth. "Bad" ozone, one of the six principal pollutants set by EPA that occurs closest to the surface of the Earth, often within 6 miles, also called *ground–level*

ozone or *troposphere ozone.* Ground–level ozone is "created by chemical reactions between oxides of nitrogen (Nox) and volatile organic compounds (VOC) in the presence of sunlight"[1]. Ground–level ozone concentrations are often high, annually in summer and daily in late morning and early afternoon. Ground–level ozone's absorption in the ultraviolet spectrum is approximately 250 nanometers (nm). Ozone is measured by comparing the degree of UV light absorption through a flow cell with ozone–free air.[2]. Ozone can also be measured by differential optical absorption spectroscopy (DOAS) instrumentation.

The US EPA sets both primary and secondary standards for ground–level ozone at 0.08 parts per million (ppm) by volume. We use parts per billion, the unit of ozone levels, instead of parts per million in order to be consistent with other studies about ozone (while noting that these units are not universally acceptable since Europe defines "billion" differently than North America, for example). Only one–hour measured ground–level ozone concentrations are considered in this thesis and so we have the primary and secondary standards[3] of 80 ppb.

Ground–level ozone concentrations are measurements of a space–time process in space–time fields where the "true" ozone levels change over time and monitoring locations. Uncertainty occurs between the measurements and the things to be measured, that is, the "true" ozone levels. We consider all these features in a Bayesian framework, a framework chosen for its great flexibility. More specifically, in this thesis, we consider fully hierarchical Bayesian models for the prediction of the ground–level ozone concentrations using AQS database or a simulated database as our application that helps

---

[1]See the following link at US EPA: http://www.epa.gov/air/ozonepollution/basic.html.

[2]See the link at: http://www.epa.qld.gov.au/environmental_management/air/air_quality_monitoring/air_pollutants/ozone/.

[3]According to EPA at http://www.epa.gov/air/ozonepollution/standards.html:

- Primary standards are the limits set "to protect public health, including the health of 'sensitive' populations such as asthmatics, children, and the elderly."

- Secondary standards are the limits set "to protect public welfare, including protection against visibility impairment, damage to animals, crops, vegetation, and buildings."

us assess our models while being of great substantive importance.

1.2 Literature Review

Multivariate models for vectors of pollutants over networks of monitoring sites prove to be much more powerful than their univariate counterparts. (See Gelfand et al. (2005) for one recent approach to the development of such models using a linear coregionalization method within a dynamic linear modeling (DLM) framework to predict the univariate and multivariate responses in space–time domains.) By combining all the information from different pollutants at multiple locations, we borrow strength and gain a better understanding of the levels of pollution at these sites. Furthermore, we obtain more accurate predictors of the pollutants at ungauged sites (unmonitored sites). Moreover, such models enable us to accommodate other site specific responses in our multivariate framework. For instance, we know that temperature affects ground–level ozone concentrations; ozone levels tend to be higher in summer than winter. Thus temperature can be incorporated directly in the model even though it would usually be regarded merely as a covariate.

Another feature of space–time fields receiving increased attention in recent literature is nonstationarity. (See Definition 2.2.2 in Chapter 2.) Nonstationarity can be due to the correlation (or covariance) varying with different site–features or heterogeneity in pollutant levels. Higdon (1998) proposes a process convolution approach to define a nonstationary process using basis function expansions. Following that, a dynamic process convolution method is proposed by Calder (2004). Calder and Cressie (2007) review various types of convolution–based models for spatial data, in which they cite Fuentes's work (Fuentes, 2002). Fuentes (2002) models the nonstationary process through the convolution method by defining that process to be a mixture of stationary processes at small subregion. Sampson and Guttorp (1992) propose a deformation approach for capturing process non-stationary while Damian et al. (2002) offer a Bayesian version of it. These covariance models fit in well with the Bayesian hierarchical models proposed by Brown

et al. (1994) who define and implement the generalized inverted Wishart distribution particularly to construct covariance models for patterned data such as those that exhibit a monotone (staircase) pattern.

We use a DLM approach to deal with the problem of nonstationarity, amongst other things. In particular, we propose a class of fully hierarchical Bayesian models that accounts for parameter uncertainty and addresses the curse of dimensionality in spatio–temporal modelling. In fact, one of the greatest difficulties arising in implementing the above approaches for spatial–temporal fields is the computational burden and inefficiency, especially in high–dimensional systems with irregularly located monitoring stations as well as sparse data. Approaches to tackling this problem include Bayesian kriging approach by Wikle and Cressie (1999), the process convolution method by Hidgon (1998), and the spatial dynamic factor approach by Lopes et al. (2007). As an extension of Wikle and Cressie, we investigate a fully Bayesian approach to construct the local principal spatial patterns through an EOF approach.

1.3   Introduction to the Thesis

My thesis contains four main themes. The first implements a version of the DLM model proposed by Huerta et al. (2004) to model an AQS database for the ozone concentrations at one cluster of 10 monitoring stations over the entire summer of 1995. This implementation, along with the background knowledge for the thesis is introduced in Chapters 2 and 3. In Chapter 2, we first review the definition and properties of spatio–temporal processes and of Kalman filtering and smoothing methods in Gaussian DLM framework. We then introduce the AQS database (1995) and an implementation of an exploratory data analysis (EDA) based on that database. Finally, we demonstrate use of MCMC algorithms for spatial interpolation and temporal prediction at ungauged sites. Following the theory presented in Chapter 2, we implement the DLM to model the hourly ozone concentrations based on the database in Chapter 3. We use the software developed for this thesis, *GDLM.1.0*, to complete this implementation. We also present potential

problems in applying this method to ground–level ozone concentrations at that space–time domain. More discussions on this implementation of the DLM can be referred to Dou et al. (2007).

Computational inefficiency proves a critical problem in using the AQS database (1995) for ozone studies – the DLM is just not scalable to large space–time domains as our work will show. Thus we explore use of an alternative, the BSP (Bayesian spatial prediction after prefiltering) or Le–Zidek approach that we implement for another AQS database (2000) to spatially interpolate and temporally predict the ground–level ozone concentrations. In Chapter 4, we first introduce related literature work and the AQS database for the Chicago area. We then demonstrate the methodology of the BSP approach. We show the existence of the spatial leakage problem (Le and Zidek, 1999) in the DLM framework, a newly result of this thesis. We summarize the results on spatial interpolation and compare them with the results using the DLM at the end of this chapter.

Le & Zidek (2006, p.131–183) suggest that a different modelling approach would be needed for the temporal prediction of univariate and multivariate responses in spatio–temporal fields. However in Chapter 5, we show how to do this with a further modelling step so that the BSP can in fact be made to yield one–day–ahead temporal forecasts for ground–level ozone concentrations. The temporal prediction results using the BSP approach are then compared with that of the DLM and another alternative called "NAIVE". We summarize these comparisons at the end of this chapter, and conclude on the advantages and disadvantages of the BSP and DLM approaches.

Because of enormous computational time savings of the BSP over the DLM approach, we chose to extend and refine the BSP, in particular, to incorporate both broad and fine scale spatial variations, while incorporating autocorrelation in the time series at each of the sites that of interest. In Chapter 6, we first show that the way EOFs ([Nancy] empirical orthogonal functions) are traditionally computed may be misleading when the time series data autocorrelated in a simulation study. [Nancy: The EOF method is used here for purpose of an extension into a fully Bayesian framework because of its widely and intensively usage in scientific community.] Assuming

a known temporal covariance function, we show the corrected EOF method better capturing main spatial patterns than the classical one. Since this covariance will always be unknown and so uncertain in practice, we propose a Bayesian EOF method to represent that uncertainty. In the second simulation study, we compare both the classical and corrected EOFs with the true EOFs, assuming a known and separable spatial–temporal covariance function. However we leave the implementation of the Bayesian EOF method. That implementation can use the MCMC algorithms already developed in this chapter.

Finally, the flexible, general structure of the DLM allows us to integrate the BSP approach into the DLM framework. Chapter 7 proposes a unified Bayesian spatio–temporal model for univariate and multivariate responses. Using this new model, we can decompose data variations into three components: long–term spatio–temporal; short–term principal spatial; and short–term spatio–temporal components. The short–term spatio–temporal components can be modelled as an BSP term. This very general, flexible model accounts for temporal correlation in the data and so allows us to update the information on those parameters as new data come in. To maintain computational speed, the Bayesian EOF has been implemented in this model to capture the local–term principal spatial patterns in the detrended spatio–temporal residual fields. We show the model's generality and flexibility by investigating that some well–known related models turn out to be special cases of ours. Those related models include the DLM by Huerta et al. (2004), Wikle and Cressie (1999), Gelfand et al. (2005), and of course, the BSP approach itself from Le and Zidek (1992). Our model also incorporates certain computational efficiencies from theoretical results we obtain in this thesis. In particular, we develop the MCMC algorithm to draw samples from the joint posterior distribution of model parameters. We can implement our model using this algorithm. However that implementation will be left to future work.

Finally in Chapter 8, we discuss future work flowing from the work of this thesis as well as possible directions for their solutions.

Chapter 2

Dynamic Linear Modelling

2.1 Introduction

We are particularly interested in the study for ozone concentrations, due to its importance for the environment. Without the shielding layer of stratosphere ozone, ultraviolet (UV) radiation would harm life on Earth. The increasing incidences of skin cancers, for example, may be linked to a thinning of Earth's ozone level. On the other hand, exceedingly high tropospheric ozone levels may cause some other diseases, for instance, eye irritation and cardiovascular diseases. We wish to study the ozone levels to more completely understand ozone today to better predict them in the future.

Ozone concentrations are a spatio–temporal field, that is, the response variable is observed across monitoring stations, which can be fixed or varied as time changing, over some time periods. There are many approaches to modelling the spatio–temporal data. We have a particular interest in dynamic linear modelling because modelling the time series process is based on "classes of dynamic models," which is often defined as "sequences of sets of models." The term, dynamic, is defined as the changes in the process "due to the passage of time as a fundamental motive force" and the DLMs (dynamic linear models) when the normality is assumed.

This chapter starts with an introduction to some basic notions in space–time theory, some elementary notions and results on the DLM and a review on Kalman filter and smoother processes in Sections 2.2–2.4.

The theme of this chapter centers on a DLM approach to modelling the ozone concentrations in the Air Quality System (AQS)[4] database. The DLM structure (proposed by Huerta et al. (2004)) is specified by an illustrative

---

[4]AQS originally called AIRS, the terminology we use hereafter.

example in Section 2.4 through some exploratory data analysis (EDA) and used throughout the following sections in this chapter. Theoretical results and algorithms on the DLM are represented in Sections 2.6 and 2.7. The MCMC sampling scheme is outlined in Section 2.6.1. The forward–filtering–backward–sampling (FFBS) method is demonstrated in Section 2.6.2 to estimate the state parameters in the DLM. Moreover, we outline the MCMC sampling scheme to obtain samples for other model parameters from their posterior conditional distributions with a Metropolis–Hasting step, and the theoretical results for prediction and interpolation at ungauged sites from their predictive posterior distributions in Section 2.7.

2.2 Space–time Process

Let $\mathcal{D}$ be the region of monitoring stations for study. For simplicity, we can fix $\mathcal{D}$ as a finite domain to study. Suppose $Y(\mathbf{s_i}, t)$ denotes the observation at time $t \in \mathcal{R}$ and site $\mathbf{s_i} \in \mathcal{D}$, and $Z(\mathbf{s_i}, t)$ denotes the space–time process at time $t \in \mathcal{R}$ and site $\mathbf{s_i} \in \mathcal{D}$, with $t = 1, \ldots, T$, and $i = 1, \ldots, n$. There may be additional covariates available, $x(\mathbf{s_i}, t)$.

The modeling structure is given by

$$Y(\mathbf{s_i}, t) = Z(\mathbf{s_i}, t) + \epsilon(\mathbf{s_i}, t), \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, \qquad (2.1)$$

where $\epsilon(\mathbf{s_i}, t)$ is a white noise process.

The space–time process $Z(\mathbf{s_i}, t)$ can be expressed as

$$Z(\mathbf{s_i}, t) = \mu(\mathbf{s_i}, t) + w(\mathbf{s_i}, t), \qquad (2.2)$$

where $\mu(\mathbf{s_i}, t)$ is the mean process obtained from the observed covariate $x(\mathbf{s_i}, t)$ and $w(\mathbf{s_i}, t)$ is a mean $\mathbf{0}$ spatio–temporal process.

Definition 2.2.1 *The space–time covariance function is defined as*

$$C(\mathbf{s_1}, \mathbf{s_2}; t_1, t_2) = Cov[w(\mathbf{s_1}, t_1), w(\mathbf{s_2}, t_2)], \qquad (2.3)$$

where $\mathbf{s_i}$ is the ith location, and t_i is a time point, for $i = 1, 2$.

Definition 2.2.2 *The zero mean spatio–temporal process $w(\mathbf{s}, t)$ is covariance stationary if*

$$C(\mathbf{s_1}, \mathbf{s_2}; t_1, t_2) = C(\mathbf{s_1} - \mathbf{s_2}; t_1 - t_2) = C(\mathbf{h}; \eta), \qquad (2.4)$$

where $\mathbf{h} = \mathbf{s_1} - \mathbf{s_2}$ and $\eta = t_1 - t_2$.

Note that $\mathbf{h}$ in (2.4) denotes the vector distance between the two sites. If $w(\mathbf{s}, t)$ does not satisfy (2.4), it is called a nonstationary spatio–temporal process.

Definition 2.2.3 *The zero mean spatio–temporal process $w(s, t)$ is isotropic if*

$$C(\mathbf{h}; \eta) = C(\|h\|; |\eta|), \qquad (2.5)$$

which means that the covariance function depends on the separation vectors only by their length of difference, $\|h\|$ and $|\eta|$. If the spatio–temporal process is not isotropic, then it is called anisotropic.

The covariance structure for the spatio–temporal process can be simplified by assuming the separability (see Definition 2.2.4).

Definition 2.2.4 *The zero mean spatio–temporal isotropic process $w(\mathbf{s}, t)$ is separable if*

$$C(\|h\|; |\eta|) = C_s(\|h\|)C_t(|\eta|), \qquad (2.6)$$

that is, the covariance function for the spatio–temporal process can be decomposed as the product of an isotropic spatial and an isotropic temporal covariance function.

If the covariance for spatio–temporal process is not separable, it is called non–separable and the process is then nonseparable.

As Gelfand et al. (2004) mentioned, for non–stationary spatial process, the general approach is to construct a valid separable covariance function. Brown et al. (1994b) and Le et al. (1998) investigate a separable covariance structure to deal with nonstationary multivariate spatial models in a

hierarchical Bayesian framework. The advantages for this approach are: (i) it reduces the number of parameters to be estimated in the model; (ii) it provides a positive definite covariance matrix.

Because of the dynamic features of spatio–temporal data, we are particularly interested in modelling the process dynamically using a dynamic modelling approach. In the following chapter, we introduce some basic notations, along with results, from the dynamic modelling.

2.3 Spatio–temporal DLM

Suppose $\mathbf{y}_t : n \times 1$ is a vector observation, for $t = 1, 2, \ldots$. The DLM contains two equations: the observation equation and the evolution equation. The observation equation is given by

$$\mathbf{y}_t = \mathbf{F}'_t \mathbf{x}_t + \nu_t, \quad \nu_t \sim N[\mathbf{0}, \mathbf{V}_t], \tag{2.7}$$

and the evolution equation is given by

$$\mathbf{x}_t = \mathbf{G}_t \mathbf{x}_{t-1} + \omega_t, \quad \omega_t \sim N[\mathbf{0}, \mathbf{W}_t], \tag{2.8}$$

where $\mathbf{F}_t : p \times n$, $\mathbf{G}_t : p \times p$, $\mathbf{V}_t : n \times n$, and $\mathbf{W}_t : p \times p$ are known matrices. In Equation (2.7), $\mathbf{F}_t$ is called the design matrix, $\mathbf{x}_t$ is the state, or system, vector, and ν_t is the observational error. Equation (2.8) is also called the evolution, state or system equation. $\mathbf{G}_t$ is the system or state matrix and ω_t is the system, or evolution, error with evolution matrix $\mathbf{W}_t$.

The dynamic linear model is completed with the initial information for the state parameter given by

$$(\mathbf{x}_0 | \mathbf{y}_0) \quad \sim \quad N[\mathbf{m}_0, \mathbf{C}_0]. \tag{2.9}$$

2.4 Kalman Filter and Smoother

Kalman recursion, or Kalman filtering as it is sometimes referred to, is used to update and forecast the state parameters. An analogous method is de-

rived by West and Harrison (1997, Chapter 4) on updating and forecasting state parameters in the DLM framework. This result can also be derived by means of Bayesian inference. Carter and Kohn (1994) label this approach forward–filtering–backward–sampling. It is particularly powerful and will be used in the following chapter. It can, moreover, be summarized in the following theorem which will be applied to the filtering and smoothing processes.

Theorem 2.4.1 *(West and Harrison, 1997) Let $\boldsymbol{y}_{1:t} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t)$, $t \geq 1$, denote all the responses observed until time t. Under models (2.7) - (2.8), together with the initial state information in (2.9), we have for $t = 2, \ldots, T$*

(i)

$$
\begin{aligned}
(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1}, \theta) &\sim N[\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}] \\
(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}, \theta) &\sim N[\boldsymbol{a}_t, \boldsymbol{R}_t] \\
(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \theta) &\sim N[\boldsymbol{f}_t, \boldsymbol{Q}_t] \\
(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}, \theta) &\sim N[\boldsymbol{m}_t, \boldsymbol{C}_t],
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{a}_t &= \boldsymbol{G}_t \boldsymbol{m}_{t-1} & \boldsymbol{R}_t &= \boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}_t' + \boldsymbol{W}_t \\
\boldsymbol{f}_t &= \boldsymbol{F}_t' \boldsymbol{a}_t & \boldsymbol{Q}_t &= \boldsymbol{F}_t' \boldsymbol{R}_t \boldsymbol{F}_t + \boldsymbol{V}_t \\
\boldsymbol{e}_t &= \boldsymbol{y}_t - \boldsymbol{f}_t & \boldsymbol{A}_t &= \boldsymbol{R}_t \boldsymbol{F}_t \boldsymbol{Q}_t^{-1} \\
\boldsymbol{m}_t &= \boldsymbol{a}_t + \boldsymbol{A}_t \boldsymbol{e}_t & \boldsymbol{C}_t &= \boldsymbol{R}_t - \boldsymbol{A}_t \boldsymbol{Q}_t \boldsymbol{A}_t'.
\end{aligned}
$$

(ii) Let $\boldsymbol{B}_t = \boldsymbol{C}_t \boldsymbol{G}_{t+1}' \boldsymbol{R}_{t+1}^{-1}$. For $0 \leq k \leq T - 1$,

$$
(\boldsymbol{x}_{T-k}|\boldsymbol{y}_{1:T}, \theta) \quad \sim \quad N[\boldsymbol{a}_T(-k), \boldsymbol{R}_T(-k)], \tag{2.10}
$$

where

$$
\begin{aligned}
\boldsymbol{a}_T(-k) &= \boldsymbol{m}_{T-k} + \boldsymbol{B}_{T-k}[\boldsymbol{a}_T(-k+1) - \boldsymbol{a}_{T-k+1}] \\
\boldsymbol{R}_T(-k) &= \boldsymbol{C}_{T-k} + \boldsymbol{B}_{T-k}[\boldsymbol{R}_T(-k+1) - \boldsymbol{R}_{T-k+1}]\boldsymbol{B}_{T-k}'
\end{aligned}
$$

with $\boldsymbol{a}_T(0) = \boldsymbol{m}_T$, $\boldsymbol{R}_T(0) = \boldsymbol{C}_T$, $\boldsymbol{a}_{T-k}(1) = \boldsymbol{a}_{T-k+1}$, and $\boldsymbol{R}_{T-k}(1) = \boldsymbol{R}_{T-k+1}$.

Note that the distributions for the smoothing and filtering processes are obtained conditionally on the parameter vector θ. In practice, it is often exceedingly difficult to obtain the posterior distribution of θ because it may not have any closed form. In this case, the MCMC method is often used to obtain samples of θ from its posterior conditional distribution. After obtaining θs, we can use Theorem 2.4.1 to get samples of state parameters from the corresponding distributions. If we can obtain all the samples of θs and state parameters, we can then do the prediction and interpolation, the general goals of the kriging method.

Next we set the DLM modelling for the Cluster 2 AQS database. This DLM modelling will be implemented in Chapter 3.

2.5 An Illustrative Example: Cluster 2 AQS Database (1995)

We consider hourly ozone concentrations in ppb measured during 1995 at 375 different monitoring stations irregularly located in the USA. Among the 375 stations, we choose three clusters of sites in close proximity with 10 monitoring stations each, for a total of 30 monitoring stations. The geographical locations of these stations are given by the latitudinal and longitudinal coordinates (see Figure 2.1).

Our goal is to construct a suitable DLM for the Cluster 2 sites based on the explanatory data analysis (EDA) as Huerta et al. (2004) suggest. The missing data are filled initially by the spatial regression method (SRM). The empirical distribution of the ozone concentrations has an asymmetric shape. So the square–root of the ozone concentrations are used as the responses due to the normality assumption of the DLM. For our study of the periodicity of the ozone dataset, we plot the Bayesian periodograms (Bretthorst, 1988) for the square–root of ozone levels for this summer of 1995 at Cluster 2 sites in Figure 2.2. We find a high peak during 1 pm to 3 pm each day for 120
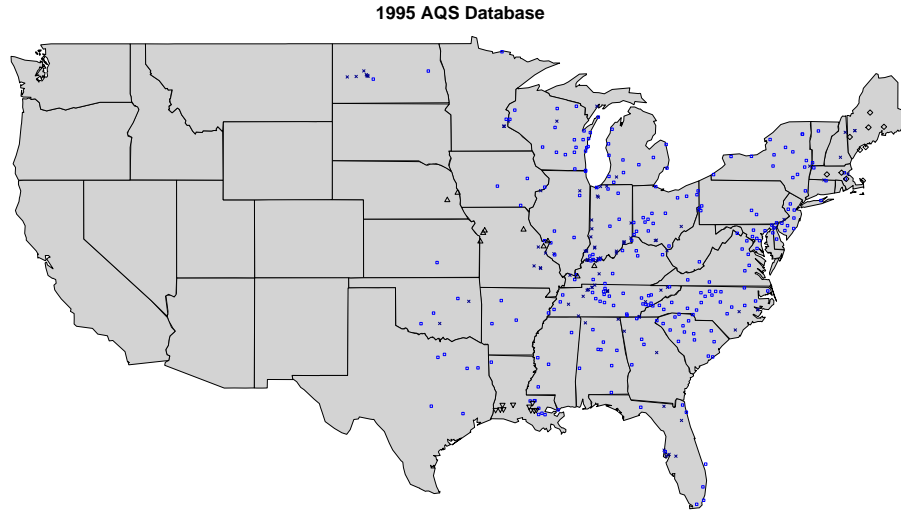
1995 AQS Database

Figure 2.1: Geographic locations for the 1995 AQS database in US map, where the latitude and longitude are measured by degrees. (Diamond = Cluster 1 sites; Upper–triangle = Cluster 2 sites; Down–triangle = Cluster 3 sites.)

days, a significant 24–hour cycle for all these gauged sites in Figure 2.2. We also find a slightly significant 12–hour cycle by plotting the periodicities which contribute more variation according to the spectrum for each station in Cluster 2 sites. We do not find any obvious weekly cycles or any nightly peaks for this database.

The DLM used in this chapter is a variation of the one proposed by Huerta et al. (2004). The state vector equation accounts for the trend and periodicity across the sites. The validity of this model for this example is assessed in Chapter 3. Given the information of other covariates, such as temperature or wind speed, better prediction of responses might be obtained using the DLM. However, due to the lack of such knowledge, we use such a variation of the DLM in Huerta et al. (2004).
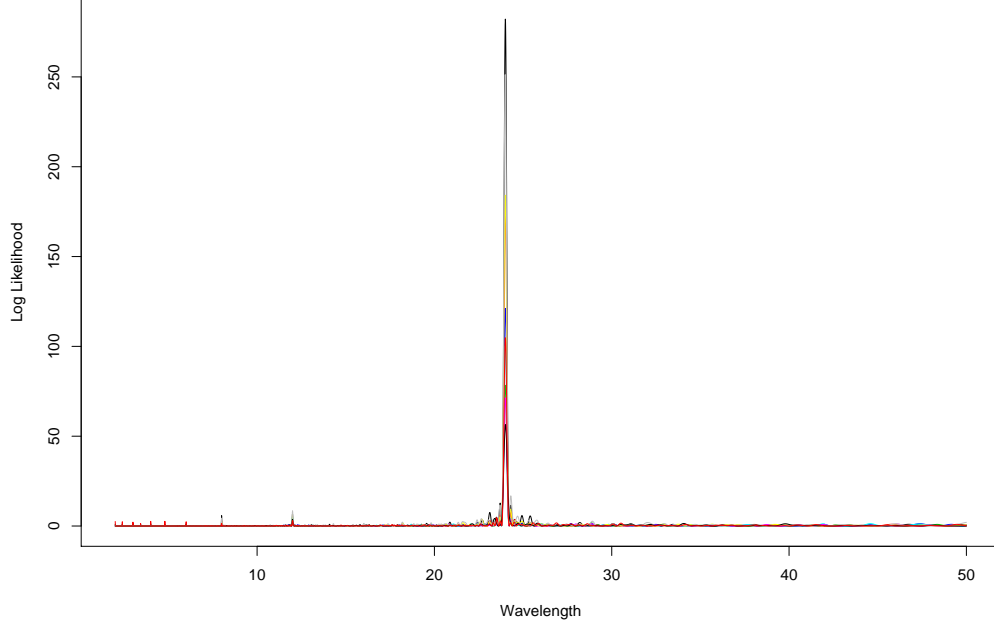
Figure 2.2: Bayesian periodogram for the square–root of hourly ozone concentrations at Cluster 2 sites in the AQS database from May 15 to September 11 (1995).

Let y_{it} denote the observed square–root of the ozone concentration, at site $\mathbf{s}_i$ and time t, with $i = 1, \ldots, n$ and $t = 1, \ldots, T$, where n presents the total number of gauged sites (that is, sites with observations) in our study and T, the total number of time points.

A variant of the state–space model for such a database is given by (Huerta et al., 2004):

$$\mathbf{y}_t = \mathbf{1}'_n \beta_t + S_{1t}(a_1)\alpha_{1t} + S_{2t}(a_2)\alpha_{2t} + \nu_t \qquad (2.11)$$

$$\beta_t = \beta_{t-1} + w_t \qquad (2.12)$$

$$\alpha_{jt} = \alpha_{j,t-1} + \omega_t^{\alpha_j}, \qquad (2.13)$$

where $\nu_t \sim N[\mathbf{0}, \sigma^2 \mathbf{V}_\lambda]$, $w_t \sim N[0, \sigma^2 \tau_y^2]$ and $\omega_t^{\alpha_j} \sim N[\mathbf{0}, \sigma^2 \tau_j^2 \mathbf{V}_{\lambda_j}]$, with $\mathbf{V}_\lambda = \exp(-\mathbf{V}/\lambda)$ and $\mathbf{V}_{\lambda_j} = \exp(-\mathbf{V}/\lambda_j)$, for $j = 1, 2$. Let $\mathbf{y}_t = (y_{1t}, \ldots, y_{nt})'$:

$n \times 1$ and $\alpha_{jt} = (\alpha_{j1t}, \ldots, \alpha_{jnt})' : n \times 1, j = 1, 2$. Here β_t denotes a canonical spatial trend and α_{jit}, a coefficient for site $\mathbf{s}_i$ at time t corresponding to a periodicity component $S_{jt}(a_j)$, where $S_{jt}(a_j) = \cos(\pi t j/12) + a_j \sin(\pi t j/12)$, for $j = 1, 2$. Note that $\mathbf{V} = (v_{ij}) : n \times n$ represents the distance matrix for the gauged sites $\mathbf{s}_1, \ldots, \mathbf{s}_n$, that is, $v_{ij} = ||\mathbf{s}_i - \mathbf{s}_j||$ for $i, j = 1, \ldots, n$. Here $||\mathbf{s}_i - \mathbf{s}_j||$ denotes the Euclidean distance between sites $\mathbf{s}_i$ and $\mathbf{s}_j$ in kilometers. Here the initial state $\mathbf{x}_0 = (\beta_0, \alpha'_{10}, \alpha'_{20})'$ has been given in (2.9).

Models (2.11)–(2.13) can also be written as in (2.7) and (2.8) by letting $\mathbf{G}_t = \mathbf{I}$, $\mathbf{V}_t = \sigma^2 \mathbf{V}_\lambda$ and $\mathbf{W}_t = \sigma^2 \mathbf{W}$, where $\mathbf{W}$ is a block diagonal matrix with diagonal entries τ_y^2, $\tau_1^2 \exp(-\mathbf{V}/\lambda_1)$, and $\tau_2^2 \exp(-\mathbf{V}/\lambda_2)$. In other words, the observation and the state equations can also be written as

$$\mathbf{y}_t = \mathbf{F}'_t \mathbf{x}_t + \nu_t \tag{2.14}$$

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \omega_t \tag{2.15}$$

where $\mathbf{x}'_t = (\beta_t, \alpha_{1t}', \alpha_{2t}')$, and $\mathbf{F}'_t$ is given by

$$
\begin{bmatrix}
1 & S_{1t}(a_1) & 0 & \ldots & 0 & S_{2t}(a_2) & 0 & \ldots & 0 \\
1 & 0 & S_{1t}(a_1) & \ldots & 0 & 0 & S_{2t}(a_2) & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \\
1 & 0 & 0 & \ldots & S_{1t}(a_1) & 0 & 0 & \ldots & S_{2t}(a_2)
\end{bmatrix}.
$$

Let $\mathbf{y}_{1:T} = (\mathbf{y}^m_{1:T}, \mathbf{y}^o_{1:T})'$, where $\mathbf{y}^m_{1:T} = (\mathbf{y}^m_1, \ldots, \mathbf{y}^m_T)$ represents all the missing values and $\mathbf{y}^o_{1:T}$, all the observed values in Cluster 2 sites for $t = 1, \ldots, T$. The model parameters are $(\lambda, \sigma^2, \mathbf{x}_{1:T}, \mathbf{y}^m_{1:T}, a_1, a_2)$, in which $\mathbf{x}_{1:T} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ are the state parameters until time T, λ the range parameter, σ^2 the variance parameter and $\mathbf{a} = (a_1, a_2)$ the phase parameters. Here we assume constant, i.e., site–invariant, phase parameters over all gauged sites due to an empirical study at each gauged site that confirms such features of spatio–temporal data in this field. Define $\gamma = (\tau_y^2, \tau_1^2, \lambda_1, \tau_2^2, \lambda_2)$ to be the vector of parameters fixed in the DLM.

The DLM is then completed with the following hyperpriors for some of

the model parameters:

$$
\begin{aligned}
\lambda &\sim IG(\alpha_\lambda, \beta_\lambda) \\
\sigma^2 &\sim IG(\alpha_{\sigma^2}, \beta_{\sigma^2}) \\
\mathbf{a} &\sim N(\mu_a^o, \boldsymbol{\Sigma}_a^o).
\end{aligned}
$$

The choice for such hyperpriors is addressed in Section 3.1.

We express the state–space model in two different ways because of our dual objectives of inference for parameters and interpolation. For simplicity, we use models (2.14)–(2.15) for the purpose of inference on the range, variance, state parameters and missingness in Sections 2.6.2–2.6.3, and use models (2.11)–(2.13) for inference about the phase parameters in Section 2.6.4 and spatial interpolation in Section 2.7.

We can see that the state–space models in (2.14)–(2.15) capture some important features of the AIRS database. It reflects the time–dependent structure of the data and captures the diurnal patterns of ozone concentrations across all the sites. Further implementation of the DLM in this database will be revisited in Chapter 3.

2.6   Algorithms for Estimating the Model Parameters

For the purpose of interpolation and prediction, one has to estimate all the unknown model parameters at each gauged site and each time point. Our goal in this section is to give the details needed to estimate the model parameters by the MCMC method and the forward–filtering–backward–sampling algorithm, developed by Carter and Kohn (1994).

In Section 2.6.1, we introduce the Metropolis–within–Gibbs method to sample from the target distribution for the model parameters given all the observations. Sections 2.6.2, 2.6.3 and 2.6.4 give details about implementing this method under the DLM. The algorithm for estimating the model parameters is then summarized in Section 2.6.5.

2.6.1 Metropolis–within–Gibbs algorithm

Consider the state space model (2.14)–(2.15). Let $\mathbf{y}_{1:T}^o = (\mathbf{y_1^o}, \ldots, \mathbf{y_T^o})$: $n \times T$ be the observation matrix at the n gauged sites until time T. Let $\mathbf{x}_{1:T} = (\mathbf{x}_1, \ldots, \mathbf{x}_T) : (2n+1) \times T$ be the state parameters at the n gauged sites until time T. For simplicity, the coordinates of γ are fixed but the problem on setting them will be addressed later in Section 3.5.

The target distribution of interest is given by $p(\lambda, \sigma^2, \mathbf{x}_{1:T}, \mathbf{y}_{1:T}^m, a_1, a_2 | \mathbf{y}_{1:T}^o)$. Since the target density does not have a closed form, direct sampling methods cannot draw samples from it. The MCMC method is a popular way to sequentially sample the parameters from their posterior distributions depending on the last value drawn and iteratively until the convergence of the chain is reached. As we can see from Huerta et al. (2004), the MCMC method can be used to obtain the posterior, predictive and interpolation results for the DLM.

A *blocking MCMC* scheme is used to sample each component iteratively from the target distribution. Three blocks are chosen as $(\lambda, \sigma^2, \mathbf{x}_{1:T})$, $\mathbf{y}_{1:T}^m$ and (a_1, a_2). There are two reasons for such blocks in applying the MCMC method. Firstly, it is natural to select blocks in which the parameters are highly correlated but relatively conditional independent between the blocks. The phase parameters are assumed independent in time and location and so less correlated with the other model parameters. Another reason derives from the fact that the full conditional posterior distribution of the phase parameters can be obtained by assuming a bivariate normal hyperprior. Details about inference are presented in Appendix A.2.

Since there is no direct way to sample from the target distribution, *Gibbs sampling* is used to sample these three blocks iteratively from their full conditional posterior distributions. In other words, we can iteratively:

(i) sample from $p(\mathbf{x}_{1:T}, \lambda, \sigma^2 | a_1, a_2, \mathbf{y}_{1:T})$,

(ii) sample from $p(\mathbf{y}_{1:T}^m | \lambda, \sigma^2, a_1, a_2, \mathbf{y}_{1:T}^o)$, and

(ii) sample from $p(a_1, a_2 | \mathbf{x}_{1:T}, \lambda, \sigma^2, \mathbf{y}_{1:T})$.

The problem is that we do not have a closed form for $p(\lambda, \sigma^2, \mathbf{x}_{1:T}|a_1, a_2,$ $\mathbf{y}_{1:T})$ either. However, the full conditional posterior distribution of $\mathbf{x}_{1:T}$ can be obtained explicitly by Kalman filtering and smoothing (Section 2.4), i.e., FFBS algorithm. Assuming an inverse Gamma hyperprior for σ^2, the conditional posterior distribution of σ^2 given the range and phase parameters also has an inverse Gamma distribution with new shape and scale parameters. Note that

$$
\begin{aligned}
p(\lambda, \sigma^2, \mathbf{x}_{1:T}|a_1, a_2, \mathbf{y}_{1:T}) & = p(\lambda|a_1, a_2, \mathbf{y}_{1:T})p(\sigma^2|\lambda, a_1, a_2, \mathbf{y}_{1:T}) \\
& \times p(\mathbf{x}_{1:T}|\lambda, \sigma^2, a_1, a_2, \mathbf{y}_{1:T}),
\end{aligned} \tag{2.16}
$$

which shows we can sample from the three conditional posterior distributions on the right–hand–side of (2.16) iteratively to obtain the samples from $p(\lambda, \sigma^2, \mathbf{x}_{1:T}|a_1, a_2, \mathbf{y}_{1:T})$. However, there is no closed form for $p(\lambda|a_1, a_2, \mathbf{y}_{1:T})$. So we must sample λ from it by a *Metropolis–Hasting* step within a Gibbs sampling cycle. This algorithm is often called *Metropolis–within–Gibbs*. Details about sampling from the joint target distribution are given in the next three subsections.

2.6.2   Sampling from $p(\lambda, \sigma^2, \mathbf{x}_{1:T}|a_1, a_2, \mathbf{y}_{1:T})$

We use the block MCMC scheme to sample $(\lambda, \sigma^2, \mathbf{x}_{1:T})$ from $p(\lambda, \sigma^2, \mathbf{x}_{1:T}|a_1,$ $a_2, \mathbf{y}_{1:T})$. Because of (2.16), ideally we can could iteratively sample λ from $p(\lambda|a_1, a_2, \mathbf{y}_{1:T})$, σ^2 from $p(\sigma^2|\lambda, a_1, a_2, \mathbf{y}_{1:T})$, and $\mathbf{x}_{1:T}$ from $p(\mathbf{x}_{1:T}|\lambda, \sigma^2, a_1,$ $a_2, \mathbf{y}_{1:T})$. However, because we do not have a closed form for that posterior density of $p(\lambda|a_1, a_2, \mathbf{y}_{1:T})$, we instead use the *Metropolis–Hasting algorithm* to sample λ, given all the observations from the following term which is proportional to its posterior density, that is,

$$
p(\lambda|a_1, a_2, \mathbf{y}_{1:T}) \propto p(\lambda) \prod_{t=1}^{T} |\mathbf{Q}_t|^{-\frac{1}{2}} \left[\beta + \frac{1}{2} \sum_{t=1}^{T} \mathbf{e}_t{}' \mathbf{Q}_t^{-1} \mathbf{e}_t \right]^{-(nT/2+\alpha)} \tag{2.17}
$$

Details are included in Appendix A.1.

Since we cannot compute the normalization constant for $p(\lambda|a_1, a_2, \mathbf{y}_{1:T})$,

the Metropolis–Hasting algorithm must be used here; it is impossible to sample λ directly from its posterior distribution. The Metropolis–Hasting algorithm yields an equilibrium distribution for the Markov chain, since computation and simulation is easier for reversible chains where the transition probabilities and stationary density of the chain satisfy the detailed balance equations. In the Metropolis–Hasting algorithm, the transition kernel is a mixed distribution for the new state of the chain: $q(.,.)$, the proposal density and $\alpha(.,.)$, the acceptance probability.

We choose the proposal density, $q(.,.)$, to be lognormal distribution, because the parameter space is bounded below by 0, the density of the Gaussian distribution making in appropriate. As Moller (2002) notes, this alternative to the random walk Metropolis considers the proposal move to be a random multiple of the current state. From the current state $\lambda^{(j-1)}(j > 1)$, the proposed move is $\lambda^* = \lambda^{(j-1)}e^Z$, where Z is drawn from a symmetric density, such as normal. In other words, at iteration j, we sample a new λ^* from this proposal distribution, centered at the previously sampled $\lambda^{(j-1)}$, with a tuning parameter τ^2 as the variance of the distribution of Z. Gammerman (2006) suggests the acceptance rate, that is, the ratio of accepted λ^* to the total number of iterations, should be around 50%. We tune τ^2 to attain that rate. If the acceptance rate is too high, for example, 70% to 100%, we then increase τ^2. Similarly, if the acceptance rate is too low, for example, 0 to 20%, we decrease τ^2 to narrow down the search area for λ^*.

The Metropolis–Hasting algorithm proceeds as follows. Given $\lambda^{(j-1)}$, where $j > 1$,

- Draw λ^* from $\text{LN}(\lambda^{(j-1)}, \tau^2)$.[5]

- Compute the acceptance probability

$$\alpha(\lambda^{(j-1)}, \lambda^*) = \min\left\{1, \frac{p(\lambda^*|\mathbf{y}_{1:T})/q(\lambda^{(j)}, \lambda^*)}{p(\lambda^{(j-1)}|\mathbf{y}_{1:T})/q(\lambda^*, \lambda^{(j-1)})}\right\}.$$

---

[5]$X \sim \text{LN}(a, b)$ means X follows a lognormal distribution. In other words, $X = \exp(Y)$ where $Y \sim N(a, b)$.

- Accept λ^* with probability $\alpha(\lambda^{(j-1)}, \lambda^*)$. In other words, sample $u \sim U[0,1]$ and let $\lambda^{(j)} = \lambda^*$ if $\lambda^* < u$ and $\lambda^{(j)} = \lambda^{(j-1)}$ otherwise.

We run this algorithm iteratively until convergence is reached.

Next we sample σ^2, given the accepted λ, a_1, a_2 and the observations. The prior for σ^2 is chosen to be an inverse gamma distribution with shape parameter α and scale parameter β. The posterior distribution for σ^2 is also an inverse gamma distribution, but with a shape parameter $\alpha + \frac{nT}{2}$ and a scale parameter $\beta + \frac{1}{2} \sum_{t=1}^{T} \mathbf{e}_t' \mathbf{Q}_t^{-1} \mathbf{e}_t$.

We now sample $\mathbf{x}_{1:T}$ given the accepted λ, σ^2, a_1, a_2 and $\mathbf{y}_{1:T}$, using the forward–filtering–backward–sampling (FFBS) method as described in Section 2.4. West and Harrison (1997) propose a general theorem for inference on the parameters in the DLM framework. For time series data, the usual method for updating and predicting is Kalman filter. We present the following theorem as the FFBS algorithm (similar to the Kalman filter algorithm) to resample the state parameters conditional on all the other parameters and observations. FFBS is used as part of a MCMC method to sample $\mathbf{x}_{1:T} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ from the smoothing distribution $p(\mathbf{x}_t | \lambda, \sigma^2, a_1, a_2, \mathbf{y}_{1:T})$. It is called FFBS because recent data are used to update the state parameters, $\mathbf{x}_t$'s, recursively for t from 1 to T, as well as to sample each element of the $\mathbf{x}_t$'s using all the information recursively for t from T to 1.

Theoretical results for models (2.14) and (2.15) are presented to give an idea of how to draw samples from the posterior distribution for $\mathbf{x}_{1:T}$. One can also obtain it from Theorem 2.4.1 in Section 2.4, by letting $\mathbf{G}_t = \mathbf{I}$, $\mathbf{V}_t = \sigma^2 \mathbf{V}_\lambda$ and $\mathbf{W}_t = \sigma^2 \mathbf{W}$.

The initial state parameter is given by

$$(\mathbf{x}_0 | \mathbf{y}_0, \theta) \sim N[\mathbf{m}_0, \sigma^2 \mathbf{C}_0], \qquad (2.18)$$

where $\mathbf{y}_0$ denotes the initial information, and $\mathbf{m}_0$ and $\mathbf{C}_0$ are known values. Later in Section 3.1, we consider how to specify them in Cluster 2 AQS database (1995). Let $\theta = (\lambda, \sigma^2, a_1, a_2, \gamma)$. Assume all the prior information has been given and θ's coordinates are mutually independent. The details on this algorithm are included in Appendix A.2.

2.6.3 Sampling from $p(\mathbf{y}_{1:T}^{m}|\lambda, \sigma^2, \mathbf{x}_{1:T}, \mathbf{y}_{1:T}^{o})$

The MCMC method has an important advantage that it enables us to fill in the missing values at each iteration, that is, to treat missing values like the model "parameters". In this way, it avoids to use *ad hoc* methods for fitted missing values, say by period means or the spatial regression method.

At any fixed time point t, after appropriately defining a scale matrix $\mathbf{R}_t$, we can rewrite the observation vector $\mathbf{y}_t$ as follows:

$$\mathbf{R}_t \mathbf{y}_t = \left(\begin{array}{c} \mathbf{y}_t^m \\ \mathbf{y}_t^o \end{array} \right),$$

where $\mathbf{y}_t^m : n_t \times 1$ denotes the missing response(s) at time t and $\mathbf{y}_t^o : (n-n_t) \times 1$ the observed response(s) at t. Notice that "o" represents for "observed" and "m" for "missing".

Let $\mathbf{R}_t = (\mathbf{e}_{n_1}, \ldots, \mathbf{e}_{n_t}, \mathbf{e}_{k_1}, \ldots, \mathbf{e}_{k_{n-n_t}})'$, where $\{\mathbf{s}_{n_j} : j = 1, \ldots, t\}$ presents the gauged sites containing missing values at time point t, $\{\mathbf{s}_{k_j} : j = 1, \ldots, n - n_t\}$ the gauged sites containing observed values at time t, for all $t = 1, \ldots, T$; and $\mathbf{e}_j$ is an $1 \times n$ vector such that $e_{jj} = 1$, $e_{jk} = 0$ if $k \neq j$, for $j \in \mathcal{Z}^+$.

We already know that

$$(\mathbf{y}_t|\lambda, \sigma^2, \mathbf{x}_t, \mathbf{a}) \quad \sim \quad N[\mathbf{F}_t'\mathbf{x}_t, \sigma^2 \exp(-\mathbf{V}/\lambda)],$$

and so $\mathbf{R}_t \mathbf{y}_t$ also has a multivariate normal distribution

$$(\mathbf{R}_t \mathbf{y}_t|\lambda, \sigma^2, \mathbf{x}_t, \mathbf{a}) \quad = \quad ((\mathbf{y}_t^m, \mathbf{y}_t^o)'|\lambda, \sigma^2, \mathbf{x}_t, \mathbf{a}) \quad \sim \quad N[\tilde{\mu}_t, \tilde{\mathbf{\Sigma}}_t],$$

where

$$\begin{aligned} \tilde{\mu}_t &= \mathbf{R}_t \mathbf{F}_t' \mathbf{x_t} \\ \tilde{\mathbf{\Sigma}}_t &= \sigma^2 \mathbf{R}_t \exp(-\mathbf{V}/\lambda) \mathbf{R}_t'. \end{aligned}$$

We can also partition $\tilde{\mu}_t$ as follows:

$$\tilde{\mu}_t \;=\; \left(\begin{array}{c} \tilde{\mu}_t^m \\ \tilde{\mu}_t^o, \end{array} \right)$$

where $\tilde{\mu}_t^m : n_t \times 1$ and $\tilde{\mu}_t^o : (n - n_t) \times 1$.

Similarly, we have

$$\tilde{\boldsymbol{\Sigma}}_t \;=\; \left(\begin{array}{cc} \tilde{\boldsymbol{\Sigma}}_t^{mm} & \tilde{\boldsymbol{\Sigma}}_t^{mo} \\ \tilde{\boldsymbol{\Sigma}}_t^{om} & \tilde{\boldsymbol{\Sigma}}_t^{oo}, \end{array} \right)$$

where $\tilde{\boldsymbol{\Sigma}}_t^{mm} : n_t \times n_t$, $\tilde{\boldsymbol{\Sigma}}_t^{mo} : n_t \times (n - n_t)$ and $\tilde{\boldsymbol{\Sigma}}_t^{oo} : (n - n_t) \times (n - n_t)$.

By a basic property of the multivariate normal distribution, we have

$$(\mathbf{y}_t^m | \lambda, \sigma^2, \mathbf{x}_t, \mathbf{a}, \mathbf{y}_t^o) \;\sim\; N[\mu_t^{**}, \boldsymbol{\Sigma}_t^{**}], \tag{2.19}$$

where

$$\mu^{**} = \tilde{\mu}_t^m + \tilde{\boldsymbol{\Sigma}}_t^{mo}(\tilde{\boldsymbol{\Sigma}}_t^{oo})^{-1}(\mathbf{y}_t^o - \tilde{\mu}_t^o), \tag{2.20}$$

and

$$\boldsymbol{\Sigma}_t^{**} = \tilde{\boldsymbol{\Sigma}}_t^{mm} - \tilde{\boldsymbol{\Sigma}}_t^{mo}(\tilde{\boldsymbol{\Sigma}}_t^{oo})^{-1}\tilde{\boldsymbol{\Sigma}}_t^{om}, \tag{2.21}$$

for $t = 1, \ldots, T$.

At each iteration, we draw the $\{\mathbf{y}_t^m\}$ from the corresponding distribution (2.19) at each time point t and then write the response variables as $\mathbf{y}_{1:T} = (\mathbf{y}_{1:T}^m, \mathbf{y}_{1:T}^o)$ without loss of generality, where $\mathbf{y}_{1:T}^m = (\mathbf{y}_1^m, \ldots, \mathbf{y}_T^m)$ and $\mathbf{y}_{1:T}^o = (\mathbf{y}_1^o, \ldots, \mathbf{y}_T^o)$.

2.6.4 Sampling from $p(a_1, a_2 | \mathbf{x}_{1:T}, \lambda, \sigma^2, \mathbf{y}_{1:T})$

We now present our method for sampling the phase parameters $\mathbf{a} = (a_1, a_2)'$ from its full conditional posterior distribution $p(\mathbf{a} | \lambda, \sigma^2, \mathbf{x}_{1:T}, \mathbf{y}_{1:T})$, using the samples for λ, σ^2 and $\mathbf{x}_{1:T}$ obtained in Sections 2.6.2–2.6.3. For simplicity, we use the notation of models (2.11)–(2.15) in this section.

We then sample the constant phase parameters conditional on all the other parameters and observations. Suppose $\mathbf{a} = (a_1, a_2)'$ has a conjugate

bivariate normal prior with mean vector $\mu^o = (\mu_{1o}, \mu_{2o})'$ and covariance matrix Σ^0. Then the posterior conditional distribution for $\mathbf{a}$ is normal with mean vector μ^* and covariance matrix Σ^*, where μ^* and Σ^* can be obtained from equations (A.13) and (A.14), respectively. This result is shown in the Appendix A.3.

We will not use a non–informative prior for $\mathbf{a}$ such as $p(\mathbf{a}) \propto 1$, because that would be problematic here. The reason is straightforward: we want to avoid cases of non–identified posterior means or posterior variances. To be more specific, assume $p(\mathbf{a}) \propto 1$. Using the same inferential approach as above, we find that the posterior conditional distribution for $\mathbf{a}$ is normal with mean vector $\mu = (\mu_1, \mu_2)'$ and covariance matrix Σ from equations (A.2) and (A.3), respectively. The elements of Σ are also given in Appendix A.3, where Σ can be singular for any $t = 12k$, where k is an integer. Hence we obtain extreme values at times $12, 24, \ldots, 2880$, which invalidates the assumption of constant phase parameters across all the time scales when we sample from their full conditional posterior distribution.

For fixed values of λ, σ^2 and $\mathbf{x}_{1:T}$, we sample $\mathbf{a}$ from $N(\mu^*, \Sigma^*)$ and then obtain the median as the estimator for $\mathbf{a}$ for each fixed iteration by exploiting the assumption that a_1 and a_2 are constant phase parameters in the models (2.14)–(2.15).

2.6.5   Summary

The MCMC methods we use here are very similar to those of Huerta et al. (2004) except that we use all the samples after the burn–in period, not just the chain corresponding to the accepted samples, as they did in their paper. That is because using only accepted Markov chains actually leads to the biases on the samples, which indeed changes the detailed balance equation of Metropolis–Hasting algorithm.

The algorithm we use in Cluster 2 AQS database is summarized as following:

```
Algorithm        The Metropolis-within-Gibbs method
```
----

1. Initialization: sample

$$
\begin{aligned}
\lambda^{(1)} &\sim IG(\alpha_\lambda, \beta_\lambda) \\
\sigma^{2(1)} &\sim IG(\alpha_{\sigma^2}, \beta_{\sigma^2}) \\
\mathbf{x}_{1:T}^{(1)} &\sim N(\mathbf{m}_0, \sigma^{2(1)}\mathbf{C}_0).
\end{aligned}
$$

2. Given the $(j-1)^{\text{th}}$ values, $\lambda^{(j-1)}$, $\sigma^{2(j-1)}$, $\mathbf{y}_{1:T}^{m\,(j-1)}$, $a_1^{(j-1)}$, $a_2^{(j-1)}$ and the observations $\mathbf{y}_{1:T}^o$:

   (1) Sample $(\lambda^{(j)}, \sigma^{2(j)}, \mathbf{x}_{1:T}^{(j)})$ from $p(\lambda, \sigma^2, \mathbf{x}_{1:T} | a_1^{(j-1)}, a_2^{(j-1)}, \mathbf{y}_{1:T}^{(j-1)})$, where

   $$
   \mathbf{y}_{1:T}^{(j-1)} = (\mathbf{y}_{1:T}^{m\,(j-1)}, \mathbf{y}_{1:T}^o).
   $$

   (i) • Generate a candidate value λ^* from a logarithm proposal distribution $q(\lambda^{(j-1)}, \lambda)$, that is, $LN(\lambda^{(j-1)}, \tau^2)$ for some suitable tuning parameter τ^2.

   • Compute the acceptance ratio $\alpha(\lambda^{(j-1)}, \lambda^*)$ where

   $$
   \alpha(\lambda^{(j-1)}, \lambda^*) = \min\left\{1, \frac{p(\lambda^* | a_1^{(j-1)}, a_2^{(j-1)}, \mathbf{y}_{1:T}^{(j-1)})\lambda^*}{p(\lambda^{(j-1)} | a_1^{(j-1)}, a_2^{(j-1)}, \mathbf{y}_{1:T}^{(j-1)})\lambda^{(j-1)}}\right\}.
   $$

   • With probability $\alpha(\lambda^{(j-1)}, \lambda^*)$ accept the candidate value and set $\lambda^{(j)} = \lambda^*$; otherwise reject and set $\lambda^{(j)} = \lambda^{(j-1)}$.

   (ii) Sample $\sigma^{2(j)}$ from $p(\sigma^2 | \lambda^{(j)}, a_1^{(j-1)}, a_2^{(j-1)}, \mathbf{y}_{1:T}^{(j-1)})$.

   (iii) Sample $\mathbf{x}_{1:T}^{(j)}$ from $p(\mathbf{x}_{1:T} | \lambda^{(j)}, \sigma^{2(j)}, a_1^{(j-1)}, a_2^{(j-1)}, \mathbf{y}_{1:T}^{(j-1)})$.

   (2) Sample $\mathbf{y}_{1:T}^{m\,(j)}$ from $p(\mathbf{y}_{1:T}^m | \lambda^{(j)}, \sigma^{2(j)}, \mathbf{x}_{1:T}^{(j)}, a_1^{(j-1)}, a_2^{(j-1)}, \mathbf{y}_{1:T}^o)$.

   (3) Sample $(a_1^{(j)}, a_2^{(j)})$ from $p((a_1, a_2) | \lambda^{(j)}, \sigma^{2(j)}, \mathbf{x}_{1:T}^{(j)}, y_{1:T}^{(j)})$, where $\mathbf{y}_{1:T}^{(j)} = (\mathbf{y}_{1:T}^{m\,(j)}, \mathbf{y}_{1:T}^o)$.

3. Repeat until convergence.

2.7 Algorithms for Interpolation and Prediction on Ungauged Sites

Our goal in this section is to interpolate the ozone concentrations at ungauged sites using the DLM and the simulated Markov chains of the model parameters (see Section 2.6). In other words, suppose $\mathbf{s}_1, \ldots, \mathbf{s}_u$ are u ungauged sites of interest within the region of Cluster 2 sites (excluding the possibility of extrapolation), the objective is to draw samples from

$$p(\mathbf{y}_{1:T}^s | \lambda, \sigma^2, \mathbf{x}_{1:T}, a_1, a_2, \mathbf{y}_{1:T}),$$

where $\mathbf{y}_{1:T}^s = (\mathbf{y}_1^s, \ldots, \mathbf{y}_T^s) : 1 \times T$ and y_t^s denotes the unobserved square–root of ozone concentrations at the ungauged site $\mathbf{s}$ and time t, for $t = 1, \ldots, T$ and $\mathbf{s} \in \{\mathbf{s}_1, \ldots, \mathbf{s}_u\}$. Let $(\alpha_{1t}^s, \alpha_{2t}^s)$ denote the unobserved state parameters at site $\mathbf{s}$ and time t. The DLM is given by

$$\mathbf{y}_t^{\text{new}} = \mathbf{1}_{n+1}'\beta_t + S_{1t}(a_1)\alpha_{1t}^{\text{new}} + S_{2t}(a_2)\alpha_{2t}^{\text{new}} + \nu_t^{\text{new}}, \quad (2.22)$$

where $\mathbf{y}_t^{\text{new}} = (y_t^s, \mathbf{y}_t')'$, $\alpha_t^{\text{new}} = (\alpha_{1t}^s, \alpha_{1t}', \alpha_{2t}^s, \alpha_{2t}')'$, and $\nu_t^{\text{new}} \sim N(\mathbf{0}, \sigma^2 \exp(-\mathbf{V}^{\text{new}}/\lambda))$.

In Section 2.7.1, we illustrate how to sample the unobserved state parameters $\{(\alpha_{1t}^s, \alpha_{2t}^s) : t = 1, \ldots, T\}$ from the corresponding conditional posterior distribution. Spatial interpolation at the ungauged site $\mathbf{s}$ is demonstrated in Section 2.7.2.

2.7.1 Sampling the unobserved state parameters

We first sample α_{jt}^s given $\alpha_{j,t-1}^s$, α_{jt} and $\alpha_{j,t-1}, j = 1, 2$. From the state equation (2.15) to α_{jt}^{new}, we know that the joint density of α_{jt}^s and α_{jt} follows a normal distribution, with covariance matrix $\sigma^2 \tau_j^2 \exp(-\mathbf{V}^{\text{new}}/\lambda_j)$, where $\mathbf{V}^{\text{new}}$ denotes the distance matrix for the unobserved station and the monitoring stations. The conditional posterior distribution,

$$p(\alpha_{jt}^s | \alpha_{j,t-1}^s, \lambda, \sigma^2, \beta_t, \alpha_{1t}, \alpha_{2t}, a_1, a_2, \mathbf{y}_{1:T}),$$

has been derived in Appendix A.4.

2.7.2   Spatial interpolation at ungauged sites

We interpolate the square–root of ozone concentration at ungauged sites by conditioning on all the other parameters and observations at gauged sites. Similarly as above, y_t^s and $\mathbf{y}_t$ are jointly normally distributed from the observation equation. The predictive conditional distribution for y_t^s, that is, $p(y_t^s | \alpha_{1t}^s, \alpha_{2t}^s, \lambda, \sigma^2, \beta_t, \alpha_{1t}, \alpha_{2t}, a_1, a_2, \mathbf{y}_{1:T})$, is given in Appendix A.4.

Chapter 3

Dynamic Linear Modelling and Its Spatial Interpolation

In Chapter 2, we illustrate the DLM settings for the Cluster 2 AQS database. Furthermore, we show the explicit MCMC method to estimate the model parameters and the predictive posterior distribution for the responses at "ungauged sites". Here we randomly select some monitoring stations in the database and treat them as "ungauged" (i.e., unmonitoring) sites for the purpose of model assessment. The theme of this chapter is the spatial interpolation using the DLM approach at the Cluster 2 AQS database. Section 3.1 revisits the Cluster 2 AIRS database. Section 3.2 shows the results of MCMC. Section 3.3 demonstrates the spatial interpolation results on the ozone study. Section 3.4 discusses the problems underlying the DLM process. Section 3.5 provides the summary and conclusions about this application of the univariate DLM.

We also develop a software, written in C and R (see Appendix B), partly solving the computational burden due to the use of MCMC algorithm in the DLM approach. This software, *GDLM.1.0,* had been successfully tested in PIMS summer school (2007). It can also be downloaded freely from http://enviro.stat.ubc.ca/dlm and the written DEMO can be directly used to illustrate the DLM in Chapter 2.

3.1 Cluster 2 AQS Dataset (1995) Revisited

Because of the importance and spatio–temporal features, ozone concentrations are of particular interest for this study. In this section, we revisit the ozone levels in Cluster 2 AQS database and study it using a variant of the

implementation of the DLM approach proposed by Huerta et al. (2004).

Within the range of Cluster 2 sites, six ungauged sites are randomly se-
lected from the available sites, that is, a non–null subset of the sites within
the range of but excluding the ten gauged ones in Cluster 2. The geograph-
ical locations of these six ungauged sites, represented by the alphabetic
letters, $A, \ldots, F$, are shown in Figure 3.1.

Cluster 2 AQS Database (1995)
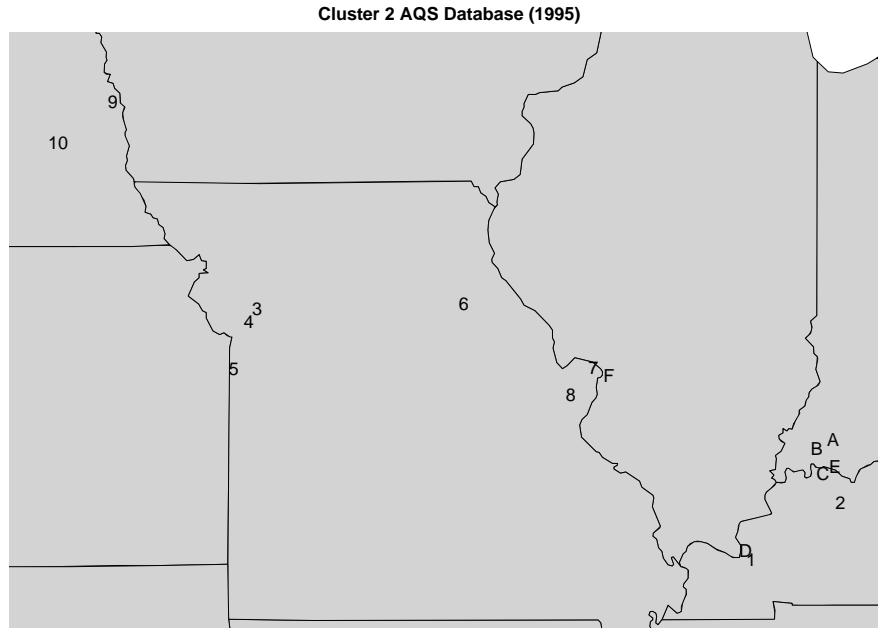
Figure 3.1: Geographical locations for the ten gauged sites in Cluster 2 and the
randomly selected six ungauged sites. (Number = Cluster 2 sites and letter =
ungauged sites.)

In Section 2.5, we have illustrated the features of ozone concentrations
tabled in the AIRS database by means of an explanatory data analysis and
constructed the DLM based on these features. In Sections 2.6.1–2.6.4, we
have demonstrated the use of FFBS and MCMC methods in the model used
in our study. Statistical inferences and results are presented in the above
subsections, where we illustrate the MCMC sampling scheme to obtain the

samples of model parameters.

For the initial values of the state parameters, hyperpriors and fixed model parameters, we use settings by Huerta et al. (2004), after confirming their suitability by preliminary investigation. We do a Markov chain simulation study and discuss the results in Section 3.2. We then interpolate the square–root of ozone concentrations at ungauged sites in Section 3.3. Issues faced in using the DLM process are discussed in Section 3.4. These issues include monitoring two Markov chain's convergence, highly autocorrelated chains for λ and σ^2, and time–varying effect of λ–σ^2 in the DLM. These problems and possible solutions are summarized in Section 3.5.

3.2   Markov Chain Simulation Study

We do a Markov chain simulation study to draw samples of the DLM's model parameters from their posterior distributions to make inference based on them.

Initial settings

As proposed by Huerta et al. (2004), we use the following initial settings for the starting values, hyperpriors and fixed model parameters in the DLM:

- The hyperprior for λ is $IG(1, 5)$ and $IG(2, 0.01)$ for σ^2. The expected value of $IG(1, 5)$ is ∞ as are both of the variances of $p(\lambda)$ and $p(\sigma^2)$. These vague priors for λ and σ^2 are selected since we do not have any prior knowledge about their distributions.

- Initially the state parameters $\mathbf{x}_0$, is assumed to be normally distributed with mean vector $\mathbf{m}_0 = (2.85, -0.75\mathbf{1}'_n, -0.08\mathbf{1}'_n)'$ and covariance matrix $\sigma_1^2 \mathbf{C}_0$, where $\sigma_1^2 \sim IG(2, 0.01)$ and $\mathbf{C}_0$ is a block diagonal matrix with diagonal entries 1, $0.011'_n$ and $0.011'_n$.

- The hyperprior for $\mathbf{a}$ is a bivariate normal distribution with mean vector $\mu^o = (2.5, 9.8)'$ and a diagonal matrix $\boldsymbol{\Sigma}^o$ with diagonal entries 0.5 and 0.5.

- The model parameters in the DLM are fixed as follows: $\tau_y^2 = 0.02$, $\tau_1^2 = 0.0002$, $\tau_2^2 = 0.0004$, $\lambda_1 = 25$, and $\lambda_2 = 25$.

We found our results to be fairly insensitive to changes in the values for μ^o, Σ^o, λ_1, and λ_2. However, it is not true for τ_y^2, τ_1^2, and τ_2^2. Further discussion on settings for these values can be found in Section 4.4 below Equation (4.24).

Monitoring the convergence of the Markov chains
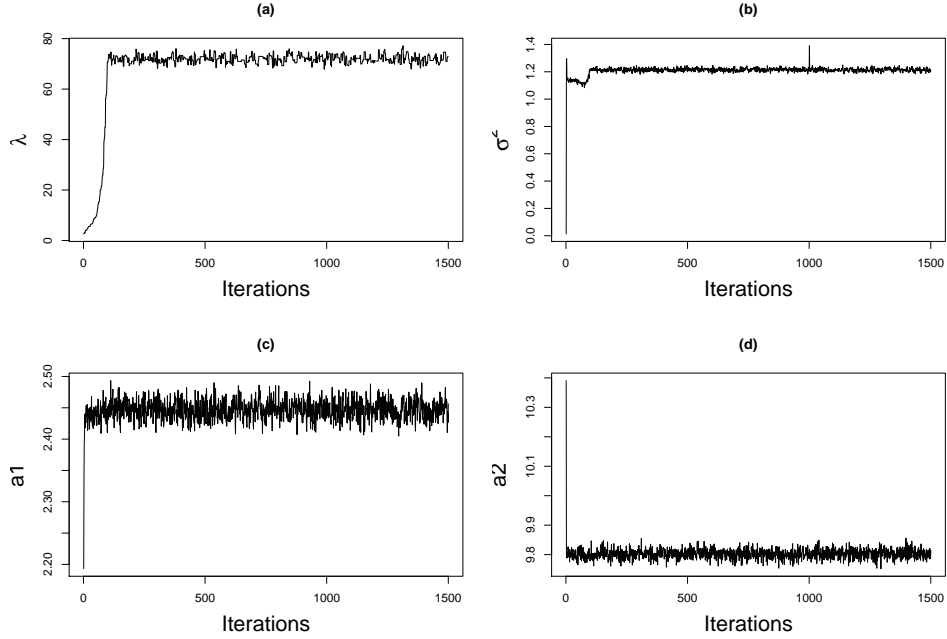
Figure 3.2: Traces of model parameters with the number of iterations of the Markov chains. The model parameters are: (a) – λ, the range parameter; (b) – σ^2, the variance parameter; (c) – a_1, the phase parameter with respect to the 24–hour periodicity; and (d) – a_2, the phase parameter with respect to the 12–hour periodicity.

Figure 3.2 shows the trace plots of model parameters λ, σ^2, a_1 and a_2 with the number of iterations of the simulated Markov chains where the total number of iterations is $4,268$. The burn–in period is chosen to be $2,269$ and

all the remaining Markov samples are collected for posterior inference. The acceptance rate is approximately 62%. We observe that the Markov chain converges after a run of less than five hundreds iterations.

Table 3.1 demonstrates the median and 95% quantile from the simulated Markov chains for the model parameters λ, σ^2, a_1 and a_2.

Quantile	λ	σ^2	a_1	a_2
2.5%	69.29	1.19	2.42	9.77
Median	71.83	1.21	2.45	9.80
97.5%	75.37	1.24	2.48	9.84

Table 3.1: Posterior summaries for λ, σ^2, a_1, and a_2.

3.3 Spatial Interpolation

Six ungauged sites are randomly selected from the available subset of sites within the range of Cluster 2 sites. Their geographical locations are shown in Figure 3.1. Our goal in this subsection is to assess the model's performance by comparing the interpolation values with the observations at these ungauged sites, that is, $A, \ldots, F$. All missing values at ungauged sites are initially filled in by the spatial regression method. The MCMC method allows us to obtain the posterior samples for the missing values from their posterior distributions. We use the observed data at ungauged sites to assess the performance of the interpolation results by the DLM.

Table 3.2 demonstrates the coverage probabilities at ungauged sites when comparing them to the corresponding credible probabilities. These coverage probabilities are calculated by counting the number of observed responses falling into the predictive intervals constructed by the DLM. It shows different levels of the predictive credibility intervals and the corresponding actual coverage probability at each of the ungauged sites based on the spatial interpolation using the DLM approach. In general, the coverage probabilities at ungauged sites are larger than the corresponding credible probability, which indicates that the error bands of the interpolation are too wide. Among these six ungauged sites, site D has the highest coverage probability at

these nominal levels. This may be due to the fact that Ungauged Site D is very close geographically to Gauged Site 1 in Cluster 2 sites. It validates our assumption that the spatial correlation is large if the pairs of sites are close together but small if they are far apart. However, these unsatisfactory coverage probabilities imply the deficiency of the DLM, which we will address in the following sections.

Nominal levels (%)	Observed coverage fraction (%)					
	A	B	C	D	E	F
95.0	94.9	96.9	96.5	99.7	96.1	98.1
90.0	91.9	93.7	93.5	99.4	93.6	96.8
80.0	84.8	88.5	88.2	97.7	89.6	94.3
70.0	78.7	83.5	83.3	94.0	85.8	90.6
60.0	73.0	78.5	77.1	89.7	81.6	86.6
50.0	65.2	71.5	70.4	85.6	76.1	81.4
40.0	55.2	61.4	61.0	79.2	67.9	74.7
30.0	42.2	47.6	47.5	69.6	54.9	64.4

Table 3.2: Comparisons between the nominal levels and actual predictive credibility interval coverage at the ungauged sites $A, \ldots, F$.

Figures 3.3–3.7 show the interpolation results at Ungauged Site D from May 14 to September 11 in 1995, where the solid lines represent the predicted median of the responses, the dashed lines represent the 95% predictive intervals for the predicted square–root of ozone concentrations and the solid dots represent the observations at this "ungauged" site.
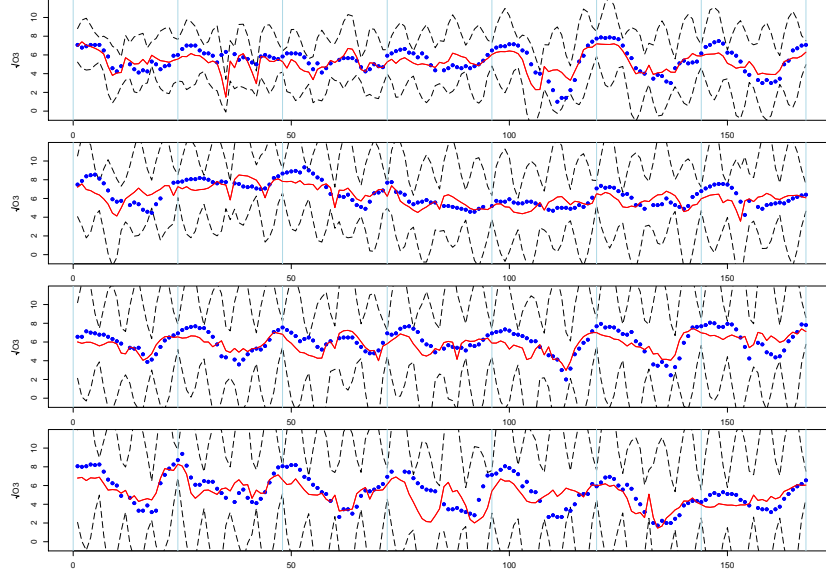
Figure 3.3: Interpolation at Ungauged Site 4 from the 1$^{\text{st}}$ week to the 4$^{\text{th}}$ week.
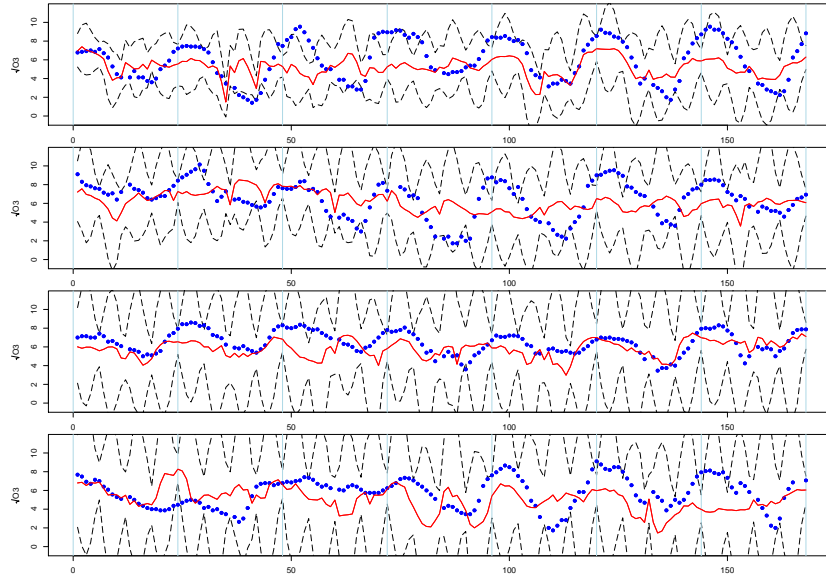
Figure 3.4: Interpolation at Ungauged Site 4 from the 5th week to the 8$^{\text{th}}$ week.
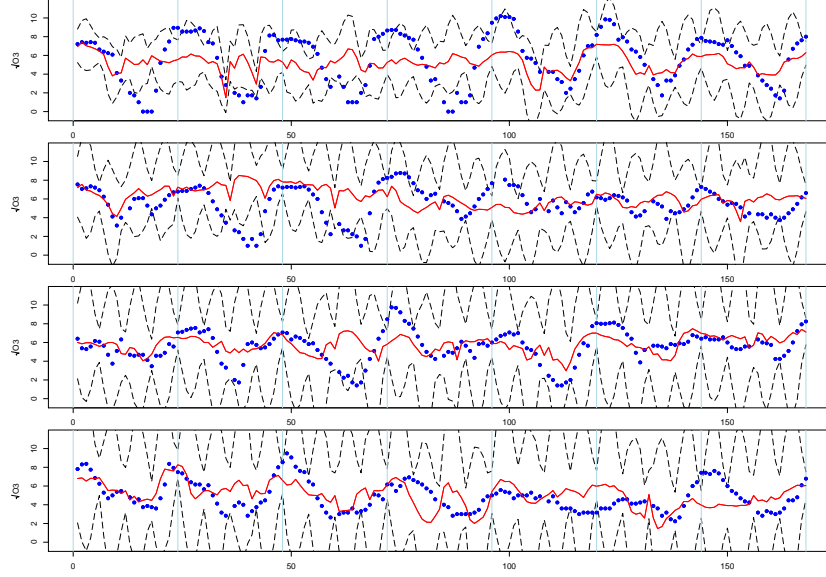
Figure 3.5: Interpolation at Ungauged Site 4 from the 9th week to the 12th week.
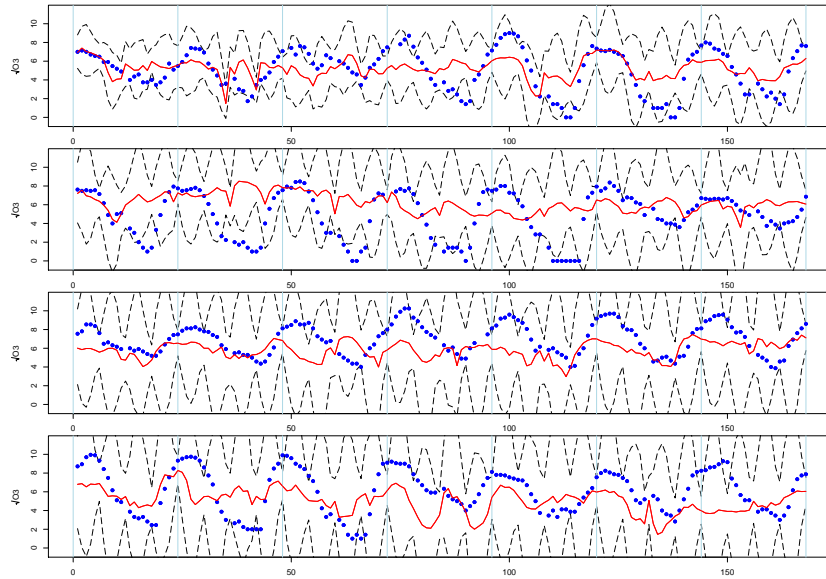
Figure 3.6: Interpolation at Ungauged Site 4 from the 13th week to the 16th week.
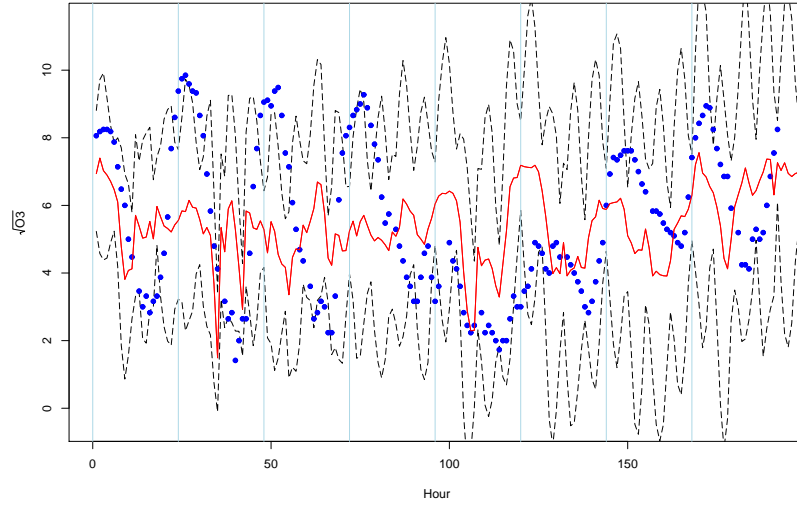
Figure 3.7: Interpolation at Ungauged Site 4 from the 17th week to the 120th day.

Table 3.2 demonstrates the coverage probabilities at each ungauged site using the data for the selected summer. That table shows Ungauged Site D having the largest coverage probability comparing to other ungauged sites. An intuitively plausible explanation: these ungauged sites are close to some gauged sites in Cluster 2 sites. To explore this possibility let us consider "friends" of ungauged sites, any gauged sites in Cluster 2 within 100 kilometers.

Table 3.3 shows the Global Circle Distance (GCD) and Pearson's correlations between these pairs of "friends". As an example, the relationship between Ungauged Site D and its "friend", Gauged Site 1, is demonstrated in Figure 3.8. From this figure, Ungauged Site D has a strong linear association with Gauged Site 1. It explains that the coverage probability at Ungauged Site D is always higher than the other sites.

Ungauged Site	Friend(s)	GCD (km)	Pearson's r
A	2	66.6	0.73
B	2	62.5	0.74
C	2	35.5	0.84
D	1	11.0	0.95
E	2	38.0	0.70
F	$(7, 8)$	$(18.6, 44.9)$	$(0.84, 0.82)$

Table 3.3: Summary of pairs of "friends" for ungauged and gauged sites.

Overall, the DLM does not predict the responses at ungauged sites very accurately. That points to problems hidden in this method and process model to which we turn in the next section.

3.4 Problems in the DLM

There are a number of problems with the current DLM. We summarize several critical issues in this section and give some suggestions about their
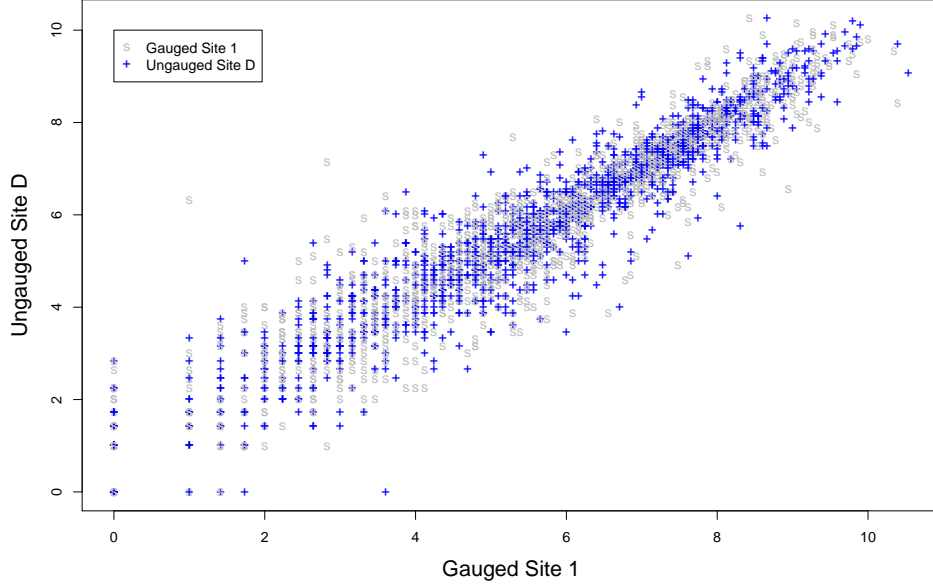
Figure 3.8: Scatterplot for the square–root of hourly ozone concentrations at Ungauged Site D and its nearly neighbour, Gauged Site 1.

resolution in the next section.

Monitoring two Markov chains' convergence

Figure 3.9 represents the trace plots of model parameters λ, σ^2, a_1, and a_2 of two chains from the initial settings in Section 3.2. These two chains seem to mix well after several hundred iterations. It suggests the Markov chains have converged.

Autocorrelation and partial autocorrelation of the Markov chains

However, we know that the autocorrelation function (ACF) is very important when considering the length of the chain needed to ensure the esti-
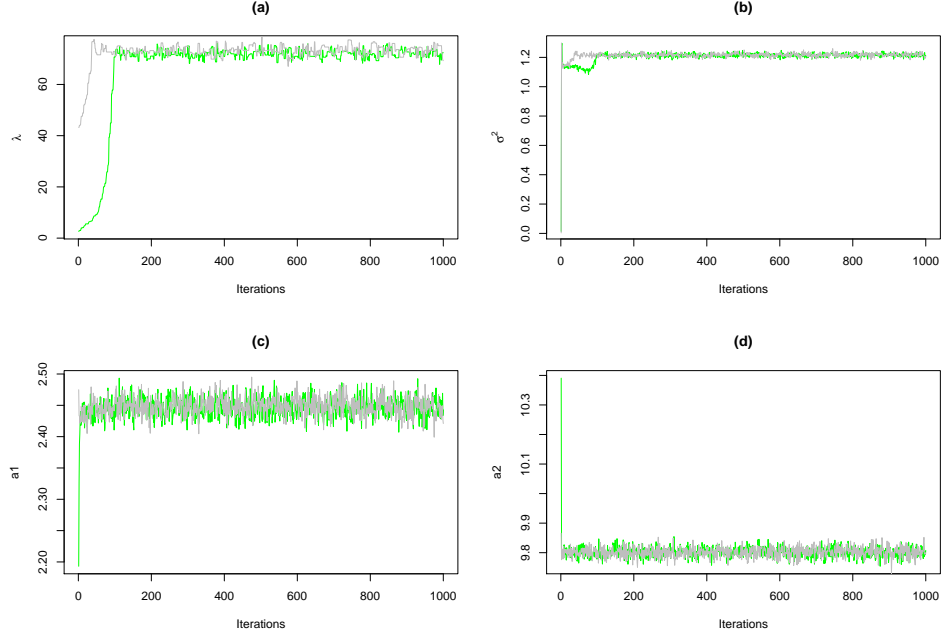
Figure 3.9: Traces of model parameters with number of iterations of the two Markov chains. The model parameters are: (a) $-\lambda$, the range parameter; (b) $-\sigma^2$, the variance parameter; (c) $-a_1$, the phase parameter with respect to the $24-$hour periodicity; and (d) $-a_2$, the phase parameter with respect to the $12-$hour periodicity.

mates having the required accuracy. In other words, a highly autocorrelated chain has to run for a long time to obtain sufficiently accurate estimates. The partial autocorrelation function (PACF) is important in assessing the Markov chain since a large value of the PACF at lag h indicates that the next value in the chain is dependent not only on the immediate past but also on the distant past.

Figure 3.10 shows the histogram as well as the ACF and PACF for the Markov chains used in Section 3.3, after a burn–in period of $1,000$. Its ACF plots show the λs to be highly autocorrelated. It indicates that the chain for λ is not mixing very well, which leads to the biased estimates in Section 3.3. A possible way to reduce the autocorrelation between these λs is to thin the Markov chain. That is, we could use every k^{th} $(k > 1, k \in \mathcal{Z}^+)$ λ

generated by the chain to give the estimates. However, due to the unduly large computational cost for thinning the Markov chain, we are forced to use the whole chain for estimation and interpolation.
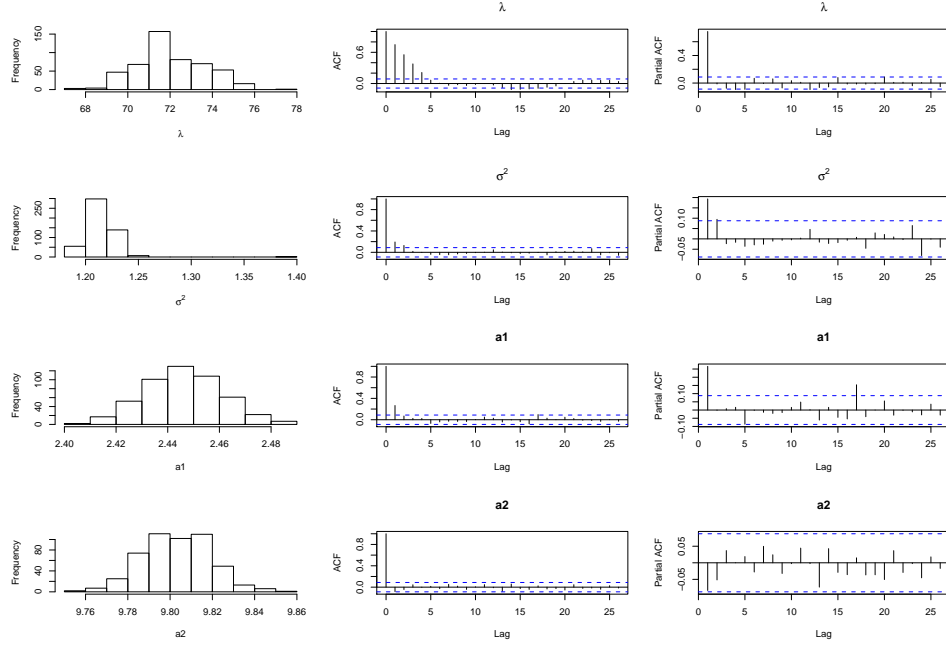
Figure 3.10: Histogram (left panel), ACF (middle panel) and PACF (right panel) of model parameters of the Markov chains after a burn–in period of $1,000$ iterations. The model parameters are: (i) first row: – λ, the range parameter; (ii) second row: – σ^2, the variance parameter; (iii) third row: – a_1, the phase parameter with respect to the 24–hour periodicity; and (iv) last row: – a_2, the phase parameter with respect to the 12–hour periodicity.

Relationship between the pairs of λ, σ^2, a_1 and a_2

The DLM assumes that priors of model parameters λ, σ^2, a_1, and a_2 are mutually uncorrelated. Figure 3.11 shows the relationship between the pairs of these parameters and specially, a weak linear association between λ and σ^2. It indicates that λ and σ^2 are actually dependent, given the observed responses.
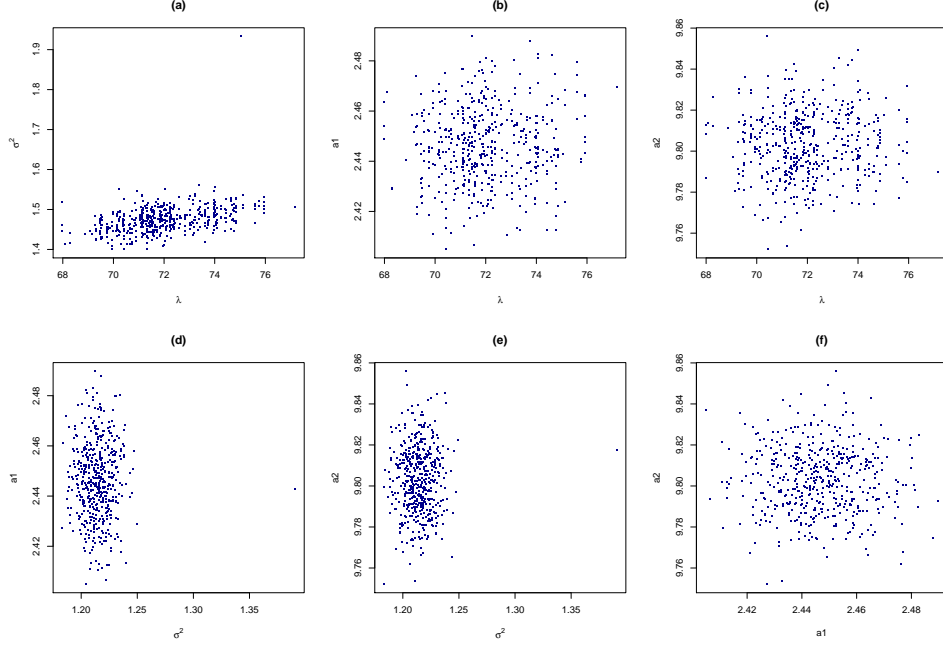
Figure 3.11: Scatterplots for model parameters' pairs: (a) λ v.s. σ^2; (b) λ v.s. a_1; (c) λ v.s. a_2; (d) σ^2 v.s. a_1; (e) σ^2 v.s. a_2; and (f) a_1 v.s. a_2.

Time varying effect of λ–σ^2: coverage probabilities versus credible probabilities

It's natural to ask whether these λs and σ^2s generated from the MCMC method are constant over all the time points, an assumption in Huerta et al.'s DLM. In other words, we want to answer questions such as: (1) Which data and time points used in the DLM might produce different estimation and interpolation result? (2) Are λ and σ^2 varying from time–to–time?

To help answer these questions, we design the following three simulation studies:

(i) Study A : Implement the DLM at ungauged sites using weekly data $(W_k : k = 1, \ldots, 17)$. Obtain the Markov chains of λ, σ^2, a_1 and a_2.

Also obtain the coverage probability at each ungauged site and each week for fixed nominal levels.

(ii) Study B : Implement the DLM at ungauged sites using all the data from week 1 to week 17 ($W_{1:17} = \{W_1, \ldots, W_{17}\}$). Estimate the model parameters and interpolate responses at ungauged sites. Furthermore, obtain the coverage probabilities at each ungauged site and each week for fixed nominal levels, using each week's data.

(iii) Study C : Fix λ_k^* at week k ($k = 1, \ldots, 17$) using the Markov chains obtained in Study A. Then use these $\lambda^* = \{\lambda_1^*, \ldots, \lambda_{17}^*\}$ in the DLM. In other words, we go through all the steps in the algorithm of Section 2.6.5 except that we use the fixed λ^* instead of generating it by a Metropolis–Hasting step. Note that we only need Gibbs sampling and MCMC blocking scheme for this study. We then compute the corresponding coverage probabilities using $W_{1:17}$ at each ungauged site and each week for fixed nominal levels.

The objective of Studies A and B is to demonstrate the effect of data and time propagation on the interpolation results. Study C aims to tell us there is a significant difference between the interpolation results obtained using the fixed λ^* and those from using the Markov samples of λs. Table 3.4 shows these fixed λ^*s in Study C.

Week	1	2	3	4	5	6	7	8	9
λ^*	54.2	178.5	83.7	405.4	86.6	59.7	199.3	144.1	322.7
Week	10	11	12	13	14	15	16	17	
λ^*	142.2	172.7	187.9	315.8	419.0	99.8	260.3	284.8	

Table 3.4: Fixed λ^* in Study C.

Figure 3.12 illustrates the MCMC estimation results for Study A. It plots the Markov chains for λ and σ^2 using weekly data. Obviously, λ and σ^2 vary from week to week, implying that the constant $\lambda - \sigma^2$ model is not tenable over a whole summer for this database.

Figures 3.13–3.19 represent the coverage probabilities of the interpolation results from these three studies. The solid line with dots represents
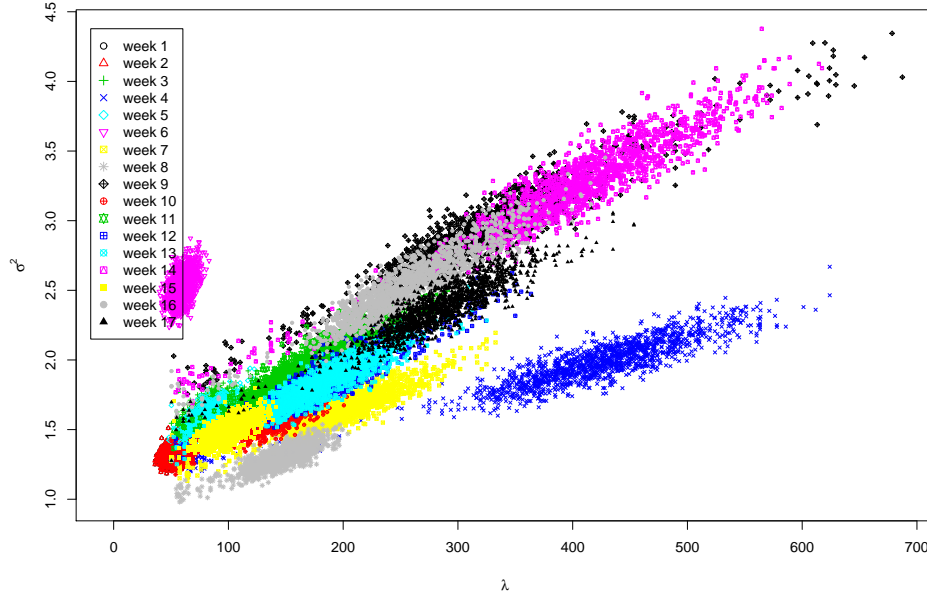
Figure 3.12: Scatterplot for λ against σ^2 given one–week–data only, constructed from MCMC samples starting from same initial values.

the results in Study A, the dot line with solid diamond for Study B, and the dashed line with stars, Study C. The interpolation results for Study B and C are very similar from these figures. In other words, using these fixed λ^*s in Table 3.4 gives us similar interpolation results as treating it to be model parameter in the standard DLM setting, pointing to a drawback with the current DLM. In fact, in some cases Study C even produces better interpolation results than Study B.

The results from Study A show that, as time increases, the interpolation results become, anomalously, more uncertain with the coverage probabilities getting larger and larger. We can interpret this as the model trying to incorporate the constant λ and σ^2 over all the time points while they actually vary with time.

Comparing these studies, sometimes the DLM gives better interpolation results when using only one week's worth of data. In any case the current

model assumption of constant λ and σ^2 is not valid in practice. Further development of the DLM is required to incorporate time–varying model parameters.
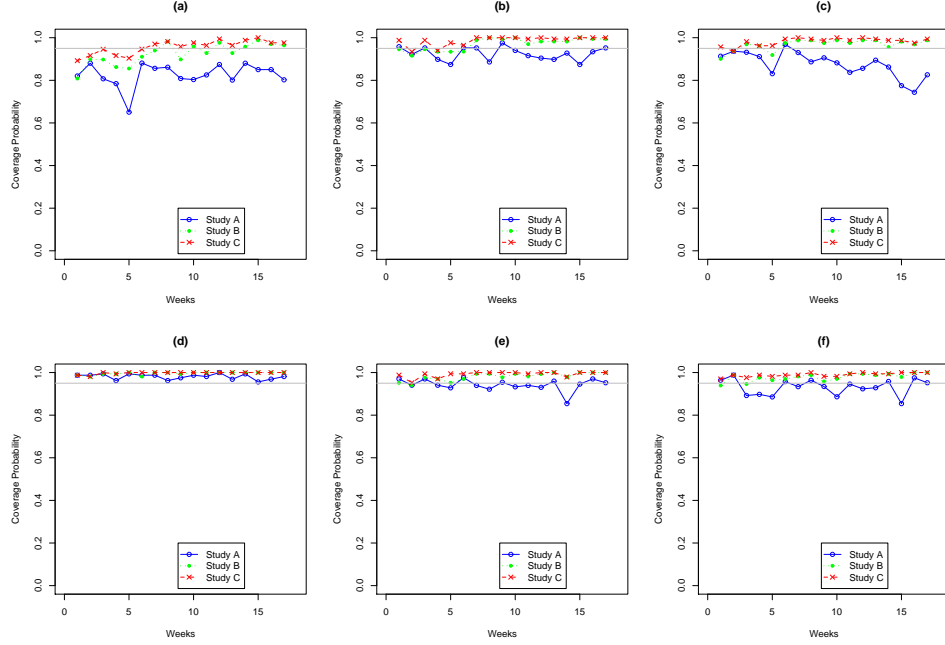
Figure 3.13: Coverage probabilities v.s. 95% nominal level for ungauged sites: (a) – site A; (b) – site B; (c) – site C; (d) – site D; (e) – site E; and (f) – site F. These coverage probabilities are computed according to *Study A*: weekly data (dot with solid line); *Study B*: $W_{1:17}$ (square with dot line); and *Study C*: $W_{1:17}$ but with fixed λ^* (star with dashed line).

3.5 Summary and Conclusion

We have implemented the DLM on Cluster 2 sites (AQS, 1995). Furthermore, we have applied a variant of Huerta et al.'s (2004) DLM and MCMC method on this database for a whole summer (whereas they considered a single week). We find deficiencies in their MCMC method which actually uses a biased estimate of λ. In practice, their model assumption of a constant λ–σ^2 seems inappropriate. Moreover, preliminary studies tell us the
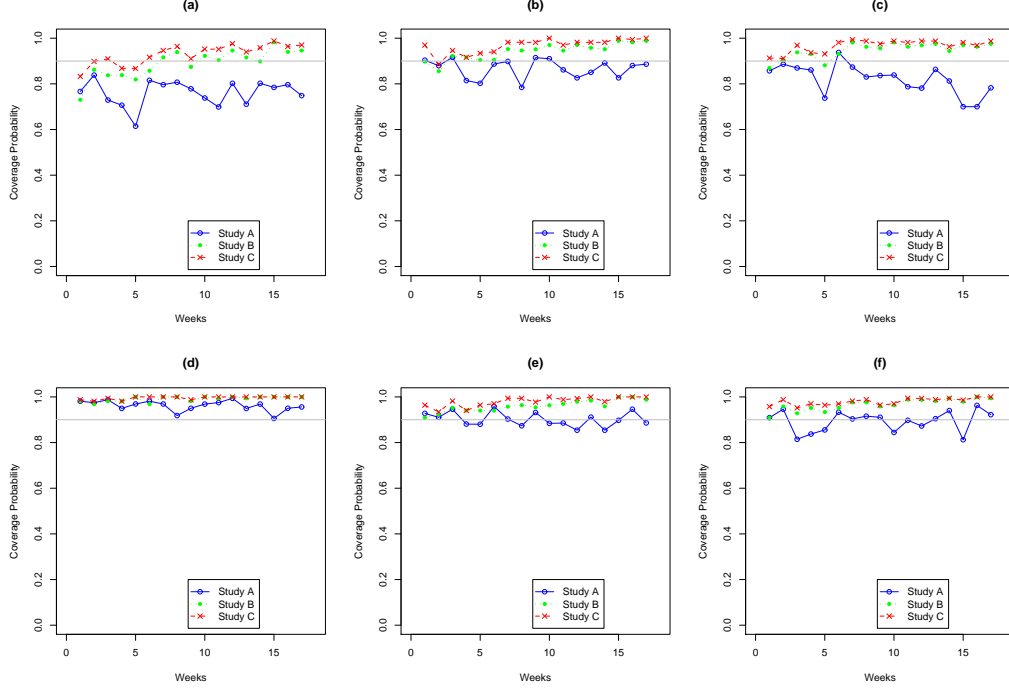
Figure 3.14: Coverage probability versus 90% nominal level for ungauged sites: (*a*) – site A; (*b*) – site B; (*c*) – site C; (*d*) – site D; (*e*) – site E; and (*f*) – site F. These coverage probabilities are computed according to *Study A*: weekly data (dot with solid line); *Study B*: $W_{1:17}$ (square with dot line); and *Study C*: $W_{1:17}$ but with fixed λ^* (star with dashed line).

sensitive choice for the values of τ_y^2, τ_1^2, and τ_2^2, indicating the inappropriate setting for these parameters to be "constant" in their model. Finally the computational cost is of concern: that cost deriving from the use of the FFBS method used in this chapter. The software for implementing the DLM, *GDLM.1.0*, has been summarized in Appendix B.

One way to tackle the interesting problem of setting τ_y^2, τ_1^2, and τ_2^2 is by setting appropriate discount factors in the discount DLM. However, we are not recommending using the composite Metropolis–Hasting algorithm to obtain the samples for $\gamma = (\lambda, \tau_y^2, \tau_1^2, \lambda_1, \tau_2^2, \lambda_2)$ from their joint posterior distribution. The reason is obvious that the computational cost is huge and it is very difficult for the Markov chains to reach its convergence after
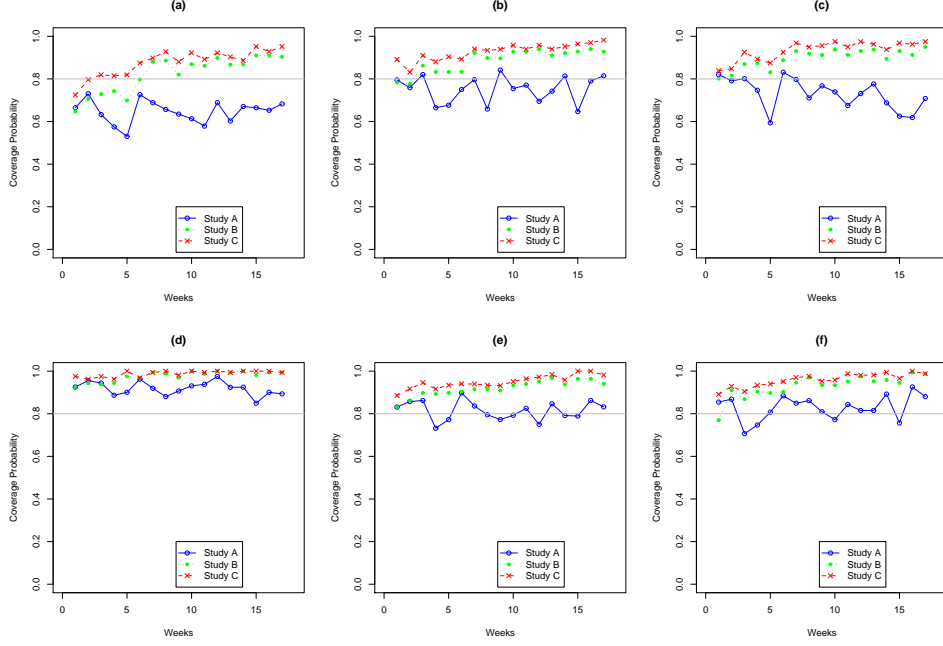
Figure 3.15: Coverage probability versus 80% nominal level for ungauged sites: (a) – site A; (b) – site B; (c) – site C; (d) – site D; (e) – site E; and (f) – site F. These coverage probabilities are computed according to *Study A*: weekly data (dot with solid line); *Study B*: $W_{1:17}$ (square with dot line); and *Study C*: $W_{1:17}$ but with fixed λ^* (star with dashed line).

certain iterations. We find the relationship between the discount factors and the model parameters: τ_y^2, τ_1^2, and τ_2^2 by a first–order polynomial dynamic model in Section 4.4.

To deal with the time–varying λ and σ^2 in the current version of the DLM, we might be able to use the discount DLM where the discount factor varies from time to time. However, this even exemplify the computational burden and so will not be recommended here. Instead of the DLM, we now propose the Le–Zidek style modelling approach (also called as BSP approach) in the next two chapters to deal with the spatial interpolation and temporal prediction in spatial–temporal fields (Le & Zidek, 1992–2006).
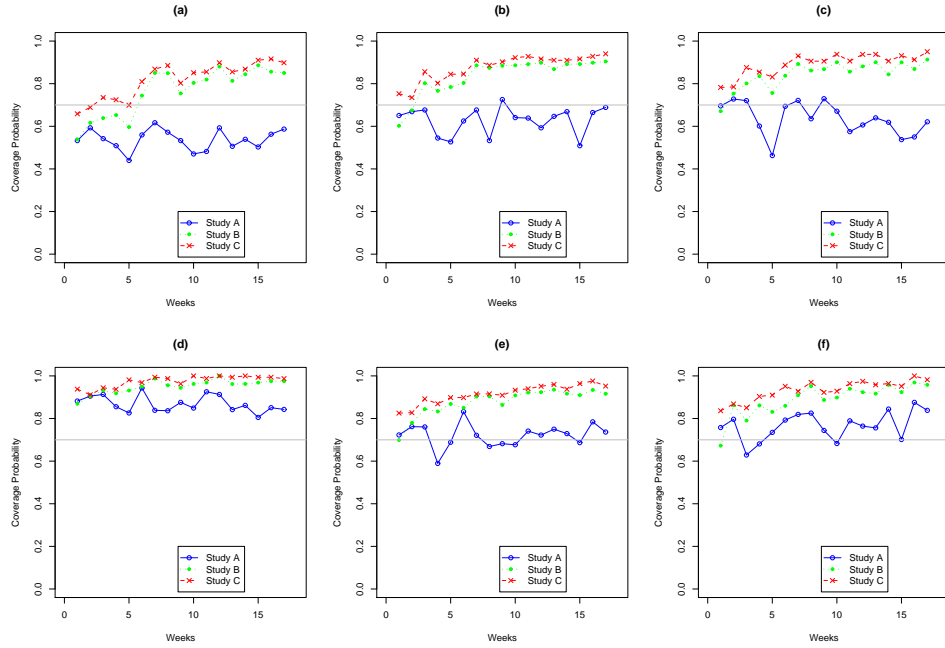
Figure 3.16: Coverage probability versus 70% nominal level for ungauged sites: (a) – site A; (b) – site B; (c) – site C; (d) – site D; (e) – site E; and (f) – site F. These coverage probabilities are computed according to *Study A*: weekly data (dot with solid line); *Study B*: $W_{1:17}$ (square with dot line); and *Study C*: $W_{1:17}$ but with fixed λ^* (star with dashed line).
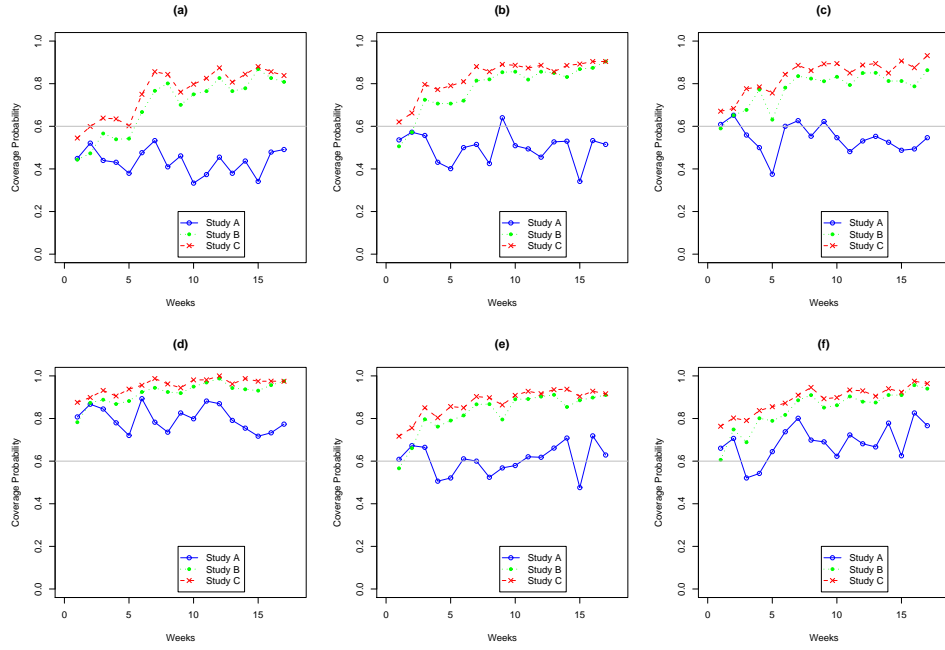
Figure 3.17: Coverage probability versus 60% nominal level for ungauged sites: (a) – site A; (b) – site B; (c) – site C; (d) – site D; (e) – site E; and (f) – site F. These coverage probabilities are computed according to *Study A*: weekly data (dot with solid line); *Study B*: $W_{1:17}$ (square with dot line); and *Study C*: $W_{1:17}$ but with fixed λ^* (star with dashed line).
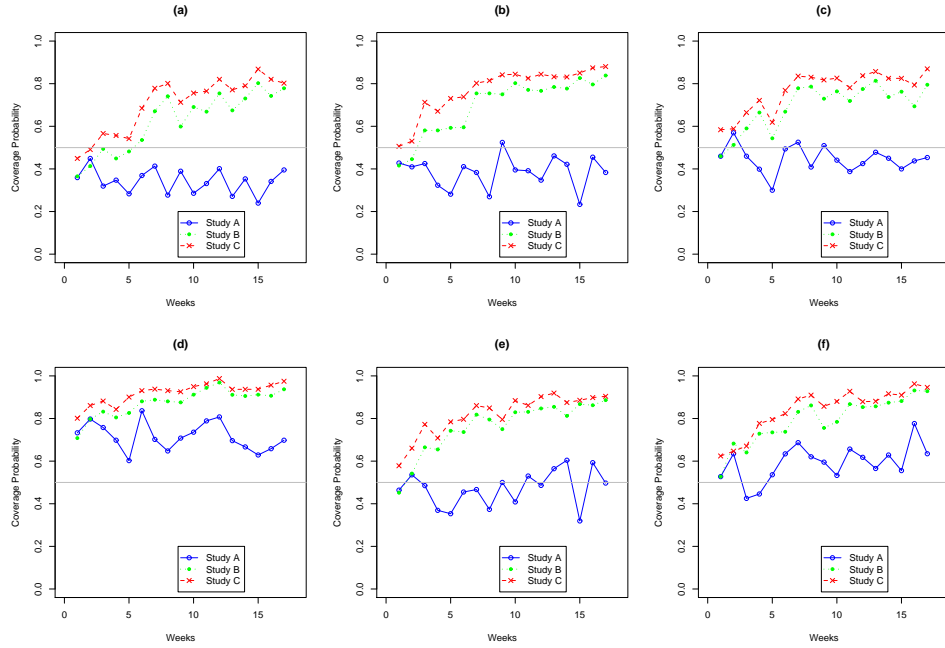
Figure 3.18: Coverage probability versus 50% nominal level for ungauged sites: (a) – site A; (b) – site B; (c) – site C; (d) – site D; (e) – site E; and (f) – site F. These coverage probabilities are computed according to *Study A*: weekly data (dot with solid line); *Study B*: $W_{1:17}$ (square with dot line); and *Study C*: $W_{1:17}$ but with fixed λ^* (star with dashed line).
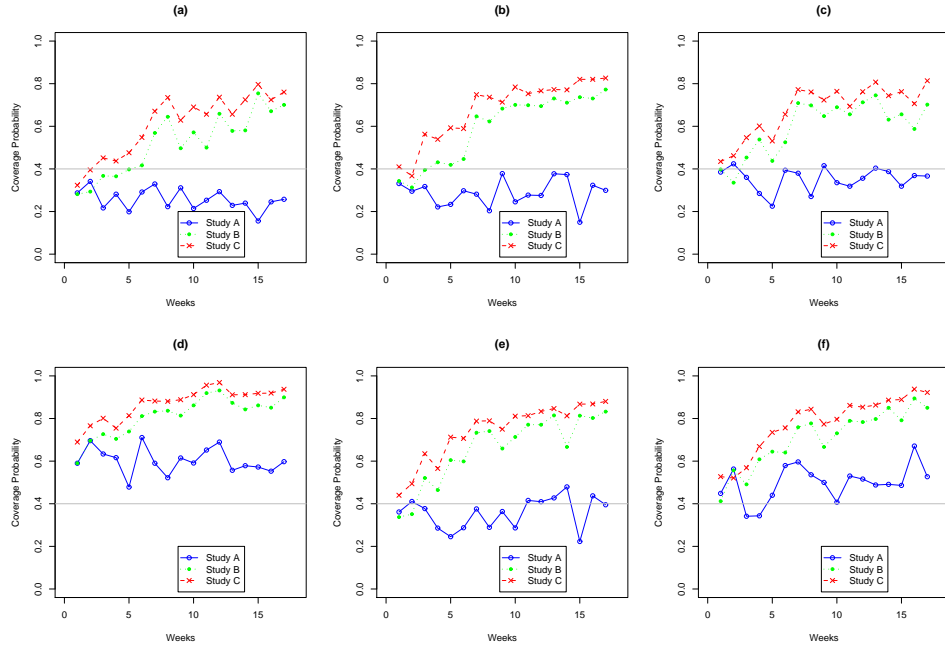
Figure 3.19: Coverage probability versus 40% nominal level for ungauged sites: (*a*) – site A; (*b*) – site B; (*c*) – site C; (*d*) – site D; (*e*) – site E; and (*f*) – site F. These coverage probabilities are computed according to *Study A*: weekly data (dot with solid line); *Study B*: $W_{1:17}$ (square with dot line); and *Study C*: $W_{1:17}$ but with fixed λ^* (star with dashed line).

Chapter 4

Multivariate Bayesian Spatial Prediction and Its Spatial Interpolation

Many approaches other than the univariate DLM (Chapter 2) have been developed to model space–time fields. Although the DLM provides a very flexible approach, it can have poor predictive performance. Moreover, this approach costs a lot of computation time and becomes impractical for large geographical subregions, for instance, for the 274 sites in the US EPA AQS database (1995). To overcome these difficulties, an alternative Bayesian hierarchical modelling approach, multivariate Bayesian spatial prediction (BSP), will be presented and its performance compared with that of the DLM.

This BSP approach appears in a series of papers, originating with Le and Zidek (1992). Section 4.1 presents that approach, its motivation, and some recent applications. Section 4.2 describes the Chicago area's hourly ozone (O_3) AQS database (2000) in this study. Section 4.3 introduces related theoretical results for the multivariate BSP approach (Le & Zidek, 2006). The multivariate BSP approach is then applied for spatial interpolation in the Chicago area's hourly ozone AQS database in Section 4.4. In Section 4.4, two other approaches, the DLM and NAIVE, are implemented for the same purposes to give comparative assessment of the performance of the various predictors. Sections 4.6 summarizes our conclusions about them.

4.1 Introduction

Alternatives to the DLM have been investigated in recent years for finding the characteristics of the mean surface of spatio–temporal processes or interpolating them at unmonitored (ungauged) sites within specified geographical subregions. Like the DLM they can deal with nonstationary processes as described in Chapter 2. In particular, the multivariate Bayesian spatial prediction (BSP) approach presented in this chapter handles nonstationary processes, using the deformation approach (Sampson & Guttorp, 1992).

The DLM treats the processes as spatially correlated time series, that is, parallel time series for a spatial process. Other approaches such as the linear model of coregionalization (LMC) method also treat data as spatially correlated time series (Gelfand et al., 2005). In contrast, the multivariate BSP treats the data as a collection of the temporally correlated spatial processes (Kyriakidis & Journel, 1999). Related work about the multivariate BSP can be seen in Le and Zidek (1992), Brown et al. (1994a, 1994b), Sun et al. (1998), Li et al. (1999), Le et al. (1997, 2001), Zidek et al. (2002), and Le and Zidek (2006).

This chapter addresses spatial interpolation in space–time fields of hourly ozone concentrations. The multivariate BSP is implemented in the Chicago area throughout the whole summer of 2000. Within this geographical region, hourly ozone concentrations are measured at 24 monitoring sites through that time span. For comparison, the DLM method (see Chapter 2) is also applied to this database.

The large geographical scale of the interpolation problem across the USA, can be handled fairly well by the multivariate BSP. Unlike what was done in Chapter 2 for the DLM, the BSP approach used in this chapter has a multivariate rather than univariate framework, although the response variable is the square–root of hourly ozone concentrations, the square–root being taken to validate the normality assumption stated in Chapter 2. We have two reasons for choosing the multivariate setting: (1) to increase precision in spatial interpolation and temporal prediction; and (2) to provide a way to reduce the *spatial correlation leakage* problem arising in the univariate case. In fact,

more accurate results can be obtained even for marginal inference based on the joint distribution model for all of the responses rather than using only a marginal distribution model, due to the greater uncertainty involved in the latter. In other words, the multivariate approach allows interpolators and predictors to "borrow strength" not only from close–by monitoring stations but also from chemical correlates (Le & Zidek, 2006, p.161). At the same time, a potential spatial correlation leakage problem, which occurs for whitened residuals (see Section 4.3) that have a nonnegligible lag 1 spatial correlation between two sites, can be side–stepped. That leakage problem seems to have been first observed by Zidek et al. (2002) while interpolating hourly PM_{10} concentrations in the Vancouver area. Though not precisely defined, the leakage happens if the cross–covariance of space and time is not negligible in the whitened residuals due for example to a failure to correctly model autocorrelation at fine temporal scales (Zidek et al., 2002). In fact, it actually provides a criterion for selecting the appropriate response vector dimension in our approach. Section 4.4 discusses appropriate initial settings for the multivariate BSP to interpolate hourly ozone concentrations. The advantages of using the multivariate approach have been described by Le and Zidek (2006), and will be reviewed in Section 4.6.

The temporal prediction problem does not at first glance seem to be embraced by the current multivariate BSP approach. The challenge arises because the multivariate response variable used in this case has to be a 24–dimensional response vector, in order to estimate the hyperscale covariance matrix among these 24 "pollutants", in reality hours, with a separability assumption on the covariance structure. Using the whole database is not appropriate because it invalidates the independence assumption on the response vectors needed. However, using a part of the database leads to a choice for the covariates, although further work on the temporal predictor has to be done. Resolutions of some above issues and theoretical results for temporal predictive distributions are presented in Section 5.1.

To assess the multivariate BSP model's performances, two other alternative approaches, the DLM and NAIVE (NAIVE*), are proposed for spatial interpolation (temporal prediction) of the Chicago area's hourly ozone con-

centration's field. These three models' performances are then compared in Section 4.4 (Sections 5.2–5.3) for the BSP, DLM and NAIVE (NAIVE*), respectively.

Computational efficiency is one major advantage of the multivariate BSP approach. The software, *EnviRo.stat*, can be freely downloaded from http://enviro.stat.ubc.ca/. Sections 4.6 and 5.5 summarize our conclusions about the implementation of this approach.

The next section describes characteristics of the Chicago area's hourly ozone field.

4.2 AQS Ozone Database (2000) for the Chicago Area

The database used in this chapter originally comes from the AQS ozone database (2000) by EPA. The hourly ground–level ozone concentrations (in ppb) for the whole summer in the Chicago area are extracted from that database. The extracted database contains 24 monitoring stations at irregularly geographical locations in this area, hourly ozone concentrations being measured at each of them. 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 4.1 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 4.2).

Figure 4.3 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 Gauged Site 3 behaves differently because
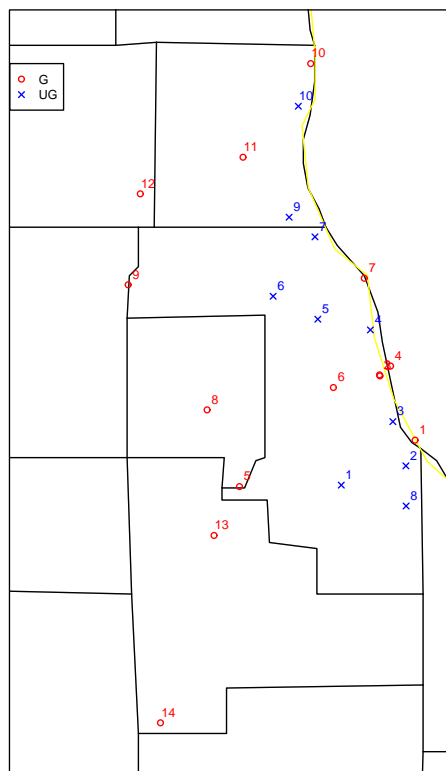
Figure 4.1: Geographical locations for the Chicago AQS database (2000), where the latitude and longitude are measured in degrees. ($\circ$ = G = gauged sites and $\times$ = UG = ungauged sites.)

of its deviation from the median for all sites and times. Figure 4.1 shows that gauged site to be near the Michigan River. However, it is unknown if the difference of the observed responses at Gauged Site 3 from the rest are due to the influence of that river or because other sites are also close to it, for example, Gauged Sites 1 and 10. One might expect that any model not taking account of this difference could lead to a poor model fit. This chapter
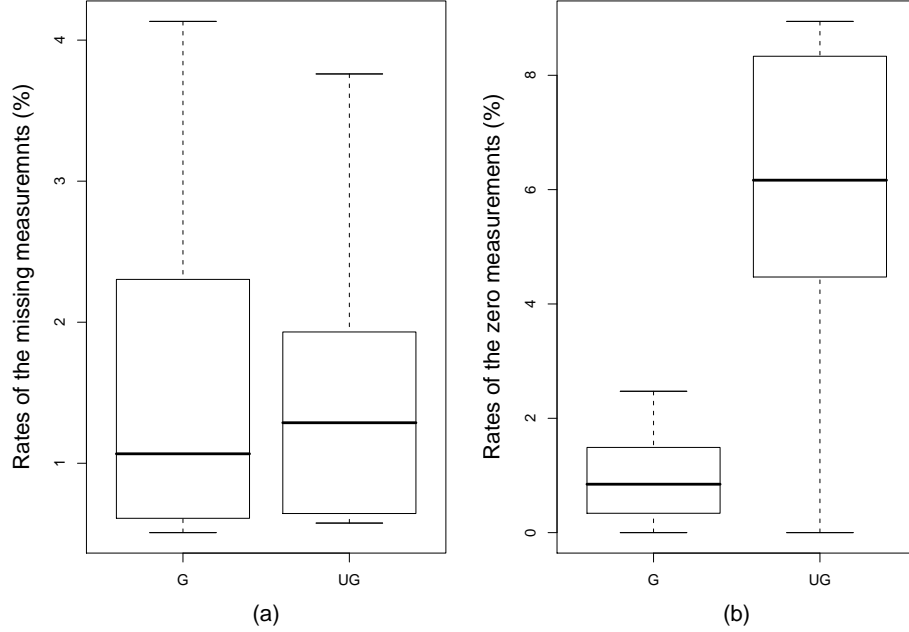
Figure 4.2: Boxplots for the rates of: (a) missing measurements; and (b) zero measurements, at 24 monitoring stations in the Chicago AQS database. (G = gauged sites and UG = ungauged sites.)

will examine this issue later when comparing interpolation (see Section 4.4) and prediction (see Section 5) of ozone concentrations' field using three different approaches: the multivariate BSP, DLM and NAIVE (NAIVE*).

To explore this database further, weekday and hourly effects are examined in Figures 4.4 and 4.5, respectively, using a simple regression method. 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. The weekday effects in Figure 4.4 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
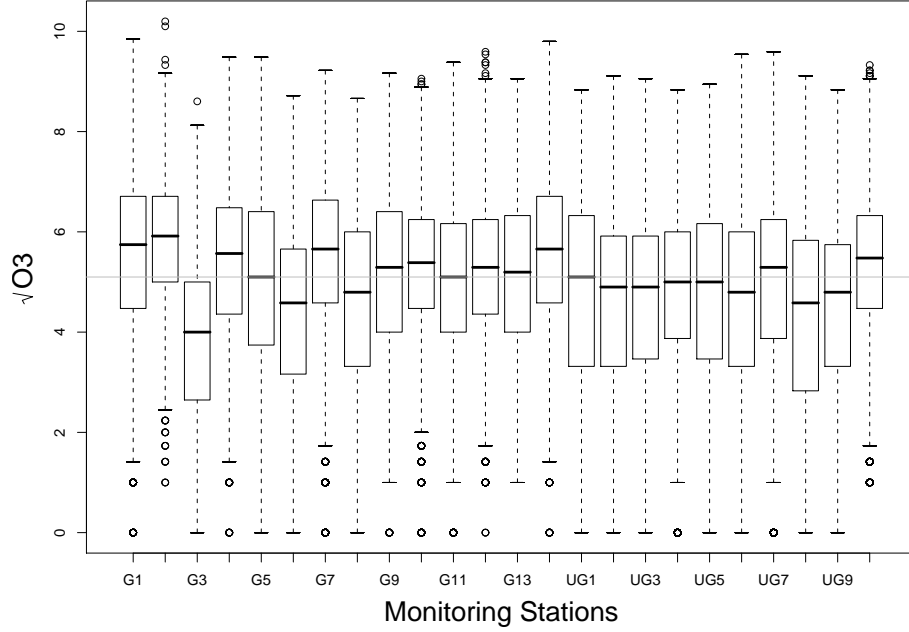
Figure 4.3: Boxplots for the square–root of hourly ozone concentrations ($\sqrt{ppb}$) at 24 monitoring stations in the Chicago AQS database. (G1 = Gauged Site 1; UG1 = Ungauged Site 1; and so on.)

and hourly effects point to constant effects for appropriate covariates in the multivariate BSP approach. Next, the corresponding model settings and methodology for the multivariate BSP is discussed in the context of the Chicago area's hourly ozone concentrations' field.

4.3 Methodology

The multivariate BSP approach puts no restriction at level one of its underlying hierarchical model structure on the covariance structure in the spatio–temporal field. Instead, the covariance function is modelled by the generalized–inverted Wishart (GIW) (Le & Zidek 1992; 2006) distribution. That approach takes account of uncertainties about the K–step staircase
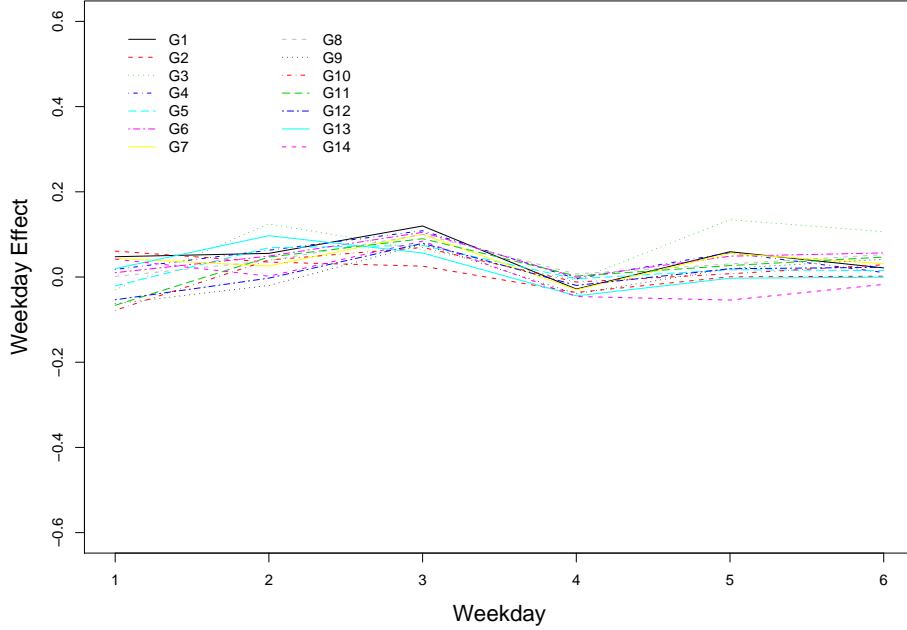
Figure 4.4: The weekday effect of the square–root of hourly ozone concentrations ($\sqrt{ppb}$) at the 14 gauged sites in the Chicago AQS database.

pattern data by allowing different degrees of freedom for different "steps" in the data array, and allows a nonstationary covariance structure in the field. In this approach, spatio–temporal multivariate responses are treated as a collection of temporally correlated spatial fields at a finite number of time points, instead of a collection of spatially correlated time series at a finite number of monitoring stations in the DLM approach (Kyriakidis & Journel, 1999). The multivariate BSP models spatio–temporal responses at two levels: at the first, the spatio–temporal random function in this field is supposed to follow a Gaussian process with the mean function depending on the covariates as well as the corresponding coefficient matrix β, and covariance function $\mathbf{\Sigma}$; at the second, the coefficient matrix β is modelled as a random matrix, following a Gaussian process with covariance function $\mathbf{\Sigma}$, having a GIW distribution. The residuals after the first level of modelling
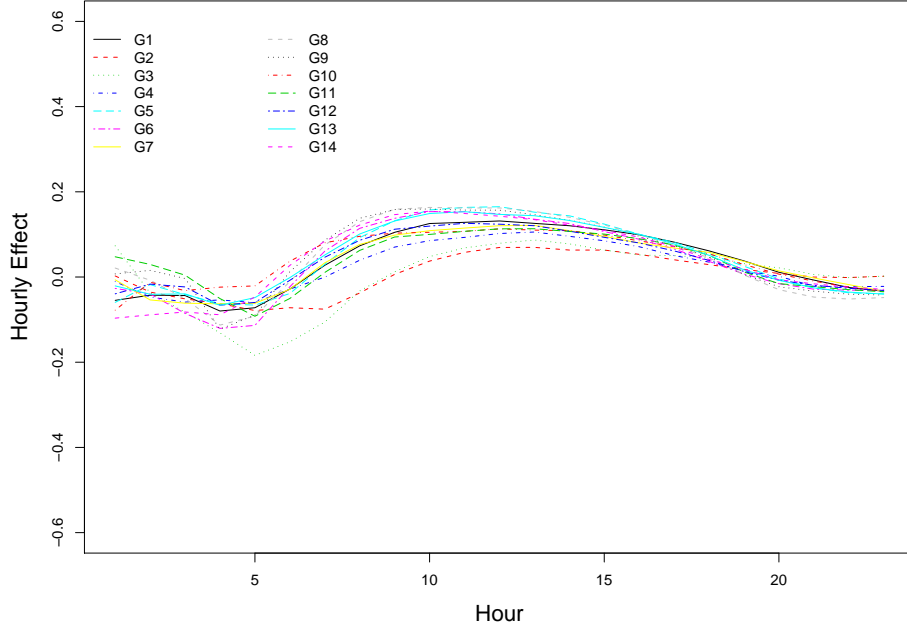
Figure 4.5: The hourly effect of the square–root of hourly ozone concentrations ($\sqrt{ppb}$) at the 14 gauged sites in the Chicago AQS database.

are called as *detrended residuals*, and those after the second level are called as *deAR'd residuals* (Zidek et al., 2002).

The multivariate BSP approach uses an autoregressive temporal structure to incorporate short–term autocorrelations and a nonstationary spatial covariance structure to deal with the nonstationary temporal–spatio processes. In particular, 2–hour–block response vectors are selected in Chicago's hourly ozone field (Section 4.2) to reduce the loss of spatial correlation leakage between the sites and allow prediction at the given hour borrowing strength from its neighbour.

Le and Zidek (2006) and Le et al. (1997) present theoretical results on the multivariate BSP model, but only the results related to the work done in this section are summarized here. Specifically, the data patterns of the multivariate BSP, that is, the *systematic missing data patterns*, are intro-

duced. Also it is demonstrated in the fully hierarchical Bayesian framework about the multivariate BSP model, as well as the conditional posterior distribution of the unobserved response variables at ungauged sites and the procedure for the hyperparameters estimation. Finally, this section presents the posterior predictive credibility ellipsoids of the unobserved responses at ungauged sites to assess the BSP's model performance (Sun et al., 1998; Le & Zidek, 2006, p.181–183). The package, *EnviRo.stat*, is used in this section (http://enviro.stat.ubc.ca). The results can be reproduced using that software since it is freely available.

Model specification

The multivariate BSP approach addresses two types of missing data patterns: *monotone missing* (i.e., *staircase pattern*) and *systematically missing* (Le et al., 1997); only the former is relevant here. The monotone missing pattern occurs when the multivariate responses are measured at different sites that start at different times. In this case, the sites can be rearranged in such a way that the data array exhibits a monotone increasing pattern over time (say a K–step staircase pattern, for $K = 1, 2, \ldots$). Within each block of the monotone missing or staircase pattern, the multivariate random vector of responses are measured from the same starting time.

Le and Zidek (2006) provide a theory that handles such data. In our application, $K = 1$ in the Chicago's hourly ozone field because the response variables are measured from the same starting time across all gauged sites. [Note: at each gauged site, the small number of missing measurements are imputed by the spatial regression method before implementing the multivariate BSP to interpolate the hourly ozone in the field. Unlike the DLM, we can obtain posterior samples for the missingness by treating them as additional parameters. Imputing the missingness by the MCMC samples might take more computational time, however.]

In this spatio–temporal field, let n denote the total number of time points, p, the total number of pollutants or species measured at each station, g, the total number of gauged sites and u, the total number of ungauged

sites. Denote the response variable at time t and gauged site g_j by

$$\mathbf{Y}_t^{[g_j]} = (Y_{t,1}^{[g_j]}, \ldots, Y_{t,p}^{[g_j]}) : 1 \times p,$$

and the response variable at time t and ungauged site u_i by

$$\mathbf{Y}_t^{[u_i]} = (Y_{t,1}^{[u_i]}, \ldots, Y_{t,p}^{[u_i]}) : 1 \times p,$$

for $t = 1, \ldots, n$, $j = 1, \ldots, g$, and $i = 1, \ldots, u$. At time t, the responses at gauged sites are coordinates of the random response vector $\mathbf{Y}_t^{[g]} = (\mathbf{Y}_t^{[g_1]}, \ldots, \mathbf{Y}_t^{[g_g]}) : 1 \times gp$ and at ungauged sites, $\mathbf{Y}_t^{[u]} = (\mathbf{Y}_t^{[u_1]}, \ldots, \mathbf{Y}_t^{[u_u]}) : 1 \times up$. The combined random response vector at time t can be written as $\mathbf{Y}_t = (\mathbf{Y}_t^{[u]}, \mathbf{Y}_t^{[g]}) : 1 \times (u+g)p$. Consequently, the matrix variate response $\mathbf{Y}$ is given by $(\mathbf{Y}_1', \ldots, \mathbf{Y}_n')' : n \times (u+g)p$. Notice that $\mathbf{Y}$ can also be written as $(\mathbf{Y}^{[u]}, \mathbf{Y}^{[g]})$, where $\mathbf{Y}^{[u]} = (\mathbf{Y}_1^{[u]'}, \ldots, \mathbf{Y}_n^{[u]'})'$, the unobserved response variables at ungauged sites, and $\mathbf{Y}^{[g]} = (\mathbf{Y}_1^{[g]'}, \ldots, \mathbf{Y}_n^{[g]'})'$, the observed response variables at gauged sites. Let $\mathbf{Z} : n \times h$ be the covariates matrix, where the total number of covariates is h. Assume that the covariates are the same across all the sites at any fixed time point. Suppose $\beta : h \times (u+g)p$ is the coefficient matrix of $\mathbf{Z}$. Assume common covariates effects across all the sites in the multivariate BSP.

The multivariate BSP model is given by

$$\begin{aligned}
\mathbf{Y}|\beta, \boldsymbol{\Sigma} &\sim N(\mathbf{Z}\beta, \mathbf{I}_n \otimes \boldsymbol{\Sigma})^2 & (4.1)\\
\beta|\boldsymbol{\Sigma}, \beta_0 &\sim N(\beta_0, \mathbf{F}^{-1} \otimes \boldsymbol{\Sigma}) & (4.2)\\
\boldsymbol{\Sigma} &\sim GIW(\boldsymbol{\Theta}, \delta), & (4.3)
\end{aligned}$$

where the covariance structure $\boldsymbol{\Sigma} : (u+g)p \times (u+g)p$ is positive definite and follows the generalized Inverted Wishart (GIW) distribution (Le & Zidek, 2006, p.158; Brown et al., 1992; Le et al., 1997; Sun et al., 1997; Le & Zidek,

1992).

The GIW distribution is defined recursively for the covariance function $\mathbf{\Sigma}$, having the K–block structure (Brown et al., 1994; Le & Zidek, 2006, p.300). The GIW can be reparameterized by the Bartlett transformation (Le & Zidek, 2006, p.302). In particular, for the case of $K = 1$, the GIW distributed $\mathbf{\Sigma}$ in (4.3) is equivalent to $(\mathbf{\Gamma}^{[u]}, \tau^{[u]}, \mathbf{\Gamma}_1)$, such that

$$\tau^{[u]}|\mathbf{\Gamma}^{[u]} \quad \sim \quad N(\tau_0^{[u]}, \mathbf{H}^{[u]} \otimes \mathbf{\Gamma}^{[u]}) \tag{4.4}$$

$$\mathbf{\Gamma}^{[u]} \quad \sim \quad IW_{up}(\delta^{[u]}, \mathbf{\Lambda}^{[u]} \otimes \mathbf{\Omega}) \tag{4.5}$$

$$\mathbf{\Gamma}_1 \quad \sim \quad IW_{gp}(\delta_1, \mathbf{\Lambda}_1 \otimes \mathbf{\Omega}), \tag{4.6}$$

where IW represents the inverted Wishart distribution. Equations (4.4)–(4.6) imply that $\tau_0^{[u]} = \mathbf{\Phi}_{gg}^{-1}\Phi_{gu}$, $\mathbf{H}^{[u]} = \mathbf{\Phi}_{gg}^{-1}$, $\delta^{[u]} = \delta$, $\mathbf{\Lambda}^{[u]} \otimes \mathbf{\Omega} = \mathbf{\Phi}_{u|g}$, $\mathbf{\Gamma}_1 = \mathbf{\Sigma}_{gg}$, $\mathbf{\Lambda}_1 \otimes \mathbf{\Omega} = \mathbf{\Phi}_{gg}$, and $\delta_1 = \delta - up$ by the properties and definitions of the IW and GIW (Le & Zidek, 2006, p.299–301). Let $\mathcal{H} = \{\mathbf{\Theta}, \delta, \mathbf{F}, \beta_0\}$, $\mathbf{\Theta} = \{\tau_0^{[u]}, \mathbf{H}^{[u]}, \mathbf{\Lambda}^{[u]}, \mathbf{\Omega}, \mathbf{\Lambda}_1\}$, and $\delta = \{\delta^{[u]}, \delta_1\}$. Hence, $\delta^{[u]} = \delta_1 + up$.

Note that sine or cose functions (show in the DLM model before representing the periodicities in spatial–temporal data) can be included in $\mathbf{Z}$, common over sites. While working with BSP, we can actually incorporate more general structure with more flexible than only using sine or cose. Moreover, we will have affordable number of parameters. To deal with the site–specific covariates, we can do it in two ways: (1) dealing it in the pre–filtering stage; or (2) treating it as a random process, and then conditioning on that response.

Given the multivariate BSP model in (4.1)–(4.3), the predictive posterior distribution of the unobserved responses at ungauged sites and the hyper-parameter estimates at the gauged and ungauged sites are given briefly in

---

[2]$\otimes$ represents the Kronecker product between two matrices such that

$$\mathbf{A}_{p \times q} \otimes \mathbf{B}_{m \times n} = \begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1q}\mathbf{B} \\ \vdots & & \vdots \\ a_{p1}\mathbf{B} & \dots & a_{pq}\mathbf{B} \end{pmatrix}_{pm \times qn} = \begin{pmatrix} b_{11}\mathbf{A} & \dots & b_{1n}\mathbf{A} \\ \vdots & & \vdots \\ b_{m1}\mathbf{A} & \dots & b_{mn}\mathbf{A} \end{pmatrix}_{pm \times qn}.$$

the next section.

Predictive distributions and hyperparameters estimation

This subsection briefly demonstrates the method used in estimating the hyperparameters in the multivariate BSP approach and the predictive posterior distribution of the multivariate response variables at ungauged sites given those estimated hyperparameters. The hyperparameters in the BSP contain the hyperparameters at gauged sites, $\mathcal{H}_g = (\mathbf{\Omega}, \mathbf{\Lambda}_1, \delta_1, \mathbf{F}, \beta_0)$, and those at ungauged sites, $\mathcal{H}_u = (\mathbf{\Lambda}^{[u]}, \mathbf{H}^{[u]}, \tau_0^{[u]}, \delta^{[u]})$. Firstly, the EM algorithm is used to estimate the hyperparameters at the gauged sites, $\mathcal{H}_g$, in the multivariate BSP. Given those estimators, the Sampson–Guttorp method is used to estimate the covariance function of the ungauged sites and the cross–covariance function between the gauged and ungauged sites. Consequently, the estimators of $\mathbf{\Gamma}^{[u]}$ and $\tau_0^{[u]}$ can be obtained from the above estimators. Finally, the predictive posterior distribution of the response variables at the ungauged sites is obtained conditional on those estimators of $\mathcal{H} = (\mathcal{H}_g, \mathcal{H}_u)$.

Note here we only estimate those hyperparameters once in the BSP approach. Not like the DLM, estimates for the model parameters have to be obtained at each iteration of the MCMC runs. This one–time–estimation greatly saves our computational time.

Suppose $\mathbf{Y} = (\mathbf{Y}^{[u]}, \mathbf{Y}^{[g]}) : (n \times up, n \times gp)$, $\beta = (\beta^{[u]}, \beta^{[g]}) : (h \times up, h \times gp)$, and $\beta_0 = (\beta_0^{[u]}, \beta_0^{[g]}) : (h \times up, h \times gp)$. Given $\mathbf{Y}^{[g]}$, the predictive distribution of $\mathbf{Y}^{[u]}$ is

$$\mathbf{Y}^{[u]} | \mathbf{Y}^{[g]}, \mathcal{H} \quad \sim \quad t_{n \times (up)}(\mu^{[u|g]}, \mathbf{\Phi}^{[u|g]} \otimes \mathbf{\Psi}^{[u|g]}, \delta - up + 1), \quad (4.7)$$

where

$$
\begin{aligned}
\mu^{[u|g]} &= \mathbf{Z}\beta_0^{[u]} + (\mathbf{Y}^{[g]} - \mathbf{Z}\beta_0^{[g]})\tau_0^{[u]} & (4.8)\\
\mathbf{\Phi}^{[u|g]} &= \mathbf{I}_n + \mathbf{Z}\mathbf{F}^{-1}\mathbf{Z}^T + (\mathbf{Y}^{[g]} - \mathbf{Z}\beta_0^{[g]})\mathbf{H}^{[u]}(\mathbf{X}^{[g]} - \mathbf{Z}\beta_0^{[g]})^T & (4.9)\\
\mathbf{\Psi}^{[u|g]} &= \frac{1}{\delta - up + 1}(\mathbf{\Lambda}^{[u]} \otimes \mathbf{\Omega}) & (4.10)
\end{aligned}
$$

(Le & Zidek, 2006, p.160–161).

At the gauged sites, the multivariate BSP can deal with the case of missing pollutants in each block. In other words, the observed responses in the matrix–variate form may include the missing columns of some pollutants at the gauged sites. Let l be the total number of unobserved responses at gauged sites, and so $l \in \{0, \ldots, gp\}$. The vector of observable responses at gauged sites $\mathbf{Y}^{[g]}$ can be partitioned into $\mathbf{Y}^{(1)} : n \times l$ and $\mathbf{Y}^{(2)} : n \times (gp - l)$, the missing and observed responses at gauged sites, respectively. Let $\mathbf{r}_j : (gp \times 1) = (r_{j,1}, \ldots, r_{j,gp})'$ be a vector such that $r_{j,j} = 1$ and $r_{j,k} = 0$ for $k \neq j$, and $k, j = 1, \ldots, gp$. Suppose $\mathbf{R}_1 = (\mathbf{r}_{i_1}, \ldots, \mathbf{r}_{i_l})$, and $\mathbf{R}_2 = (\mathbf{r}_{i_{l+1}}, \ldots, \mathbf{r}_{i_{gp}})$; thus $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2) : (gp \times gp)$, forms an orthogonal matrix. Hence, $\mathbf{Y}^{(1)} = \mathbf{Y}^{[g]}\mathbf{R}_1$ and $\mathbf{Y}^{(2)} = \mathbf{Y}^{[g]}\mathbf{R}_2$. Let

$$\mathbf{R}^T\boldsymbol{\Sigma}_{gg}\mathbf{R} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{R}_1^T\boldsymbol{\Sigma}_{gg}\mathbf{R}_1 & \mathbf{R}_1^T\boldsymbol{\Sigma}_{gg}\mathbf{R}_2 \\ \mathbf{R}_2^T\boldsymbol{\Sigma}_{gg}\mathbf{R}_1 & \mathbf{R}_2^T\boldsymbol{\Sigma}_{gg}\mathbf{R}_2 \end{pmatrix}$$

and

$$\boldsymbol{\Psi}_{gg} = \mathbf{R}^T\boldsymbol{\Phi}_{gg}\mathbf{R} = \begin{pmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{R}_1^T\boldsymbol{\Phi}_{gg}\mathbf{R}_1 & \mathbf{R}_1^T\boldsymbol{\Phi}_{gg}\mathbf{R}_2 \\ \mathbf{R}_2^T\boldsymbol{\Phi}_{gg}\mathbf{R}_1 & \mathbf{R}_2^T\boldsymbol{\Phi}_{gg}\mathbf{R}_2 \end{pmatrix},$$

with $\boldsymbol{\Sigma}_{11}, \boldsymbol{\Psi}_{11} : l \times l$ and $\boldsymbol{\Sigma}_{22}, \boldsymbol{\Psi}_{22} : (gp - l) \times (gp - l)$, respectively. Similarly, $\beta_0^{[g]}\mathbf{R}$ can be partitioned as $\beta_0^{[g]} = (\beta_0^{[g]}\mathbf{R}_1, \beta_0^{[g]}\mathbf{R}_2) = (\beta_{(1)}^{[g]}, \beta_{(2)}^{[g]})$.

By the properties of the multivariate t–distribution, the predictive posterior distribution of $\mathbf{Y}^{(1)}$ is given by

$$\mathbf{Y}^{(1)}|\mathbf{Y}^{(2)}, \mathcal{H} \quad \sim \quad t_{n \times h}(\mathbf{Z}\beta_{(1)}^{[g]} + (\mathbf{Y}^{(2)} - \mathbf{Z}\beta_{(2)}^{[g]})\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Psi}_{21}, \frac{1}{\delta - up - l + 1}\mathbf{P}_{1|2}$$
$$\otimes \boldsymbol{\Psi}_{1|2}, \delta - up - l + 1), \tag{4.11}$$

where

$$\mathbf{P}_{1|2} \quad = \quad \mathbf{I}_n + \mathbf{Z}\mathbf{F}^{-1}\mathbf{Z}^T + (\mathbf{Y}^{(2)} - \mathbf{Z}\beta_{(2)}^{[g]})\boldsymbol{\Psi}_{22}^{-1}(\mathbf{Y}^{(2)} - \mathbf{Z}\beta_{(2)}^{[g]})^T \tag{4.12}$$

and

$$\boldsymbol{\Psi}_{1|2} = \boldsymbol{\Psi}_{11} - \boldsymbol{\Psi}_{12}\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Psi}_{21}. \tag{4.13}$$

Furthermore, the predictive posterior distribution of $\mathbf{Y}^{[u]}$ is given as follows

$$\mathbf{Y}^{[u]}|\mathbf{Y}^{(2)}, \mathcal{H} \sim t_{n \times up}(\mathbf{Z}\beta_0^{[u]} + (\mathbf{Y}^{(2)} - \mathbf{Z}\beta_{(2)}^{[g]})\boldsymbol{\Psi}_{22}^{-1}\mathbf{R}_2^T\boldsymbol{\Phi}_{gu}, \frac{\mathbf{P}_{u|2} \otimes \boldsymbol{\Phi}_{u|g}}{\delta - up - l + 1},$$
$$\delta - up - l + 1), \tag{4.14}$$

where

$$\mathbf{P}_{u|2} = \mathbf{I}_n + \mathbf{Z}\mathbf{F}^{-1}\mathbf{Z}^T + (\mathbf{Y}^{(2)} - \mathbf{Z}\beta_{(2)}^{[g]})\boldsymbol{\Psi}_{22}^{-1}(\mathbf{Y}^{(2)} - \mathbf{Z}\beta_{(2)}^{[g]})^T \tag{4.15}$$

(Le et al., 1997).

The next subsection explores the predictive performance of the multivariate BSP approach.

Predictive performance

To assess the multivariate BSP model's performance, pointwise predictive intervals and predictive posterior credibility ellipsoids are constructed from the predictive posterior distributions in Section 4.3 (Le et al., 1997; Le & Zidek, 2006).

Le and Zidek (2006) point out that the pointwise predictive distribution of the last (or p^{th}) pollutant at each of the ungauged sites and any fixed time point t can be obtained from (4.14) by letting $up = 1$. Moreover, the pointwise predictive variance is given by

$$\text{Var}(\mathbf{Y}_t^{[u]}|\mathbf{Y}^{(2)} = \mathbf{y}^{(2)}, \mathcal{H}) = (\delta^* - 2)^{-1}P_{t|2}\boldsymbol{\Phi}_{u|2}, \tag{4.16}$$

where $\delta^* = \delta - up - l + 1$. Hence, the pointwise predictive intervals of the last pollutant at ungauged site u_i and the fixed time t is given by

$$\text{E}(\mathbf{Y}_t^{[u]}|\mathbf{Y}^{(2)} = \mathbf{y}^{(2)}, \mathcal{H}) \pm t_{\delta^*}(0.025)(\text{Var}(\mathbf{Y}_t^{[u]}|\mathbf{Y}^{(2)} = \mathbf{y}^{(2)}, \mathcal{H}))^{1/2}.$$

However, one might expect that the pointwise predictive intervals would not have good calibration properties due to the accumulated uncertainty arising from simultaneously interpolating the response variables in the spatio–temporal field. Le et al. (1997) develop the ellipsoid credible regions for simultaneously interpolating at ungauged sites, improving its calibration. Those ellipsoid credible regions are given by the following theorem that we include here for completeness.

Theorem 4.3.1 *The* $(1-\alpha)-level$ $(0 < \alpha < 1)$ *simultaneously posterior credibility region is given by*

$$\{\, \boldsymbol{Y}_t^{[u]} : (\,\boldsymbol{Y}_t^{[u]} - \hat{\boldsymbol{y}}_t^{[u]})\boldsymbol{\Phi}_{u|g}^{-1}(\,\boldsymbol{Y}_t^{[u]} - \hat{\boldsymbol{y}}_t^{[u]})^T < \frac{up}{\delta - up - l + 1}P_{t|2}F_{up,\delta-up-l+1}(1-\alpha)\},$$

where

$$\hat{\boldsymbol{y}}_t^{[u]} = \boldsymbol{Z}_t\beta_0^{[u]} + (\,\boldsymbol{Y}_t^{[g]} - \boldsymbol{Z}_t\beta_0^{[g]})\tau_0^{[u]},$$

$$P_{t|2} = 1 + \boldsymbol{Z}_t\boldsymbol{F}^{-1}\boldsymbol{Z}_t^T + (\,\boldsymbol{Y}_t^{[g]} - \boldsymbol{Z}_t\beta_0^{[g]})\boldsymbol{H}^{[u]}(\,\boldsymbol{Y}_t^{[g]} - \boldsymbol{Z}_t\beta_0^{[g]})^T,$$

and $\boldsymbol{\Phi}_{u|g}^{-1} = (\boldsymbol{\Lambda}^{[u]} \otimes \boldsymbol{\Omega})^{-1}.$

Theorem 4.3.1 is true because

$$\boldsymbol{Y}_t^{[u]}|\boldsymbol{Y}_t^{(2)},\mathcal{H} \quad \sim \quad t_{1\times up}(\hat{\mathbf{y}}_t^{[u]}, \frac{1}{\delta - up - l + 1}P_{t|2}\boldsymbol{\Phi}_{u|g}, \delta - up - l + 1),$$

$$(4.17)$$

where

$$\hat{\mathbf{y}}_t^{[u]} \quad = \quad \mathbf{Z}_t\beta_0^{[u]} + (\mathbf{Y}_t^{(2)} - \mathbf{Z}_t\beta_{(2)}^{[g]})\boldsymbol{\Psi}_{22}^{-1}\mathbf{R}_2^T\boldsymbol{\Phi}_{gu}, \qquad (4.18)$$

and

$$P_{t|2} \quad = \quad 1 + \mathbf{Z}_t\mathbf{F}^{-1}\mathbf{Z}_t^T + (\mathbf{Y}_t^{(2)} - \mathbf{Z}_t\beta_{(2)}^{[g]})\boldsymbol{\Psi}_{22}^{-1}(\mathbf{Y}_t^{(2)} - \mathbf{Z}_t\beta_{(2)}^{[g]})^T. \quad (4.19)$$

Further work by Sun et al. (1998) shows these credibility ellipsoids to be well calibrated. These credible regions are also constructed and compared with the pointwise predictive intervals when interpolating in the Chicago's

hourly ozone field, in the next section.

4.4   Spatial Interpolation

This section's main theme concerns the interpolation of ozone concentrations at those ten ungauged sites using the multivariate BSP approach for the Chicago's hourly ozone field. Moreover, two other approaches, the DLM and NAIVE, are also used in the same interpolation problem to assess the model's performance. The interpolation results using these three approaches are then compared in this section.

The multivariate BSP approach

For the multivariate BSP model given by (4.1)–(4.3), hourly ground–level ozone concentrations's field is modelled as a trend plus a *detrended spatio–temporal noise* (i.e., *detrended residuals*) at the first level of a hierarchical model; at the second level, the temporal dependence in the detrended spatio–temporal noise is modelled through an autoregressive process with the *deAR'd residuals.* These deAR'd residuals are then interpolated at the ungauged sites in the spatio–temporal field. Finally, these interpolated deAR'd residuals are imputed by taking them back to the trend model to get the interpolated values at the ungauged sites. To do this, we square the interpolated responses due to the square–root transformation we made due to the normality assumption assumed for this model.

The multivariate BSP model has been used to interpolate Vancouver's hourly PM_{10} field (Li et al., 1999; Zidek et al., 2002). A multivariate data model is created to reduce the loss of spatial correlation in the deAR'd residuals compared with their detrended counterparts, and to allow the predictions to be made on any response by borrowing strength from the close–by responses through their spatial correlations (Le et al., 1999; Zidek et al., 2002; Le & Zidek, 2006, p.277). This loss of spatial correlation has been called the *spatial correlation leakage* problem (also seen in Section 4.5) by Li et al. (1999), Zidek et al. (2002), and Le and Zidek (2006). The spatial

leakage problem occurs if nonnegligible lag–1, lag–2, etc., spatial correlations exist in the deAR'd residuals. Using those problematic deAR'd residuals to interpolate at ungauged sites, one may anticipate that the interpolators at any given ungauged site cannot borrow strength from its gauged neighbours due to the small number of spatial correlation between sites (Zidek et al., 2002; Le & Zidek, 2006). Figures 4.6 plots the spatial correlations of the detrended and deAR'd residuals between sites, respectively. The substantial loss of the spatial correlation of the deAR'd residuals suggests the use of the multivariate strategy adopted here. In particular, that strategy enables one to bypass the need for difficult fine scale autocorrelation modelling that will inevitably be inexact and hence induce the lagged spatial cross–correlations that can cause that leakage.
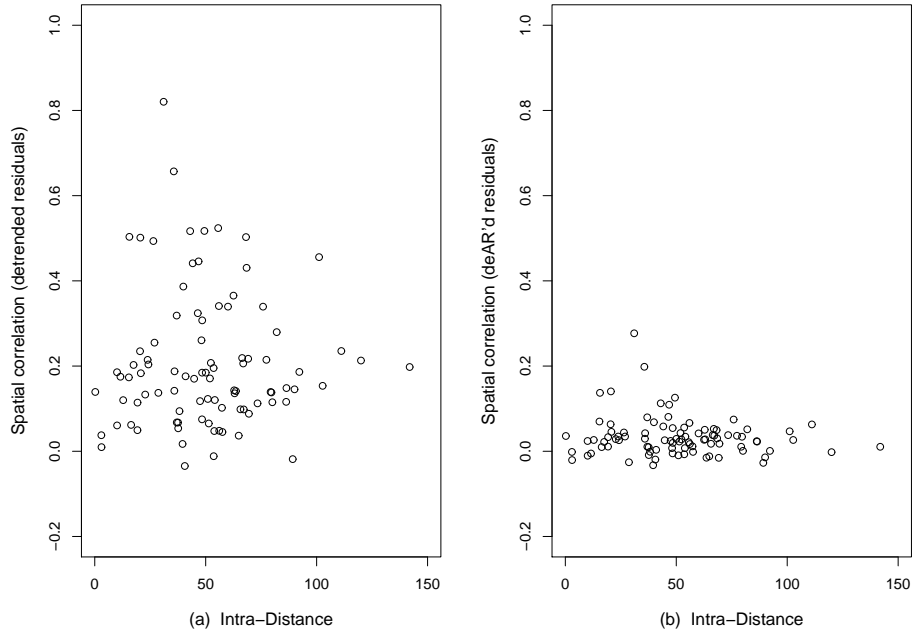
Figure 4.6: The estimated spatial correlations of: (a)–detrended residuals, and (b)–deAR'd residuals; between gauged sites.

But what kind of multivariate data settings should one select? Le et

al. (1999) and Zidek et al. (2002) argue for the use of the daily response vectors with hourly response coordinates instead of the hourly responses themselves. One of their arguments, that the spatial leakage problem is negligible for the daily responses but not for the hourly ones, is supported by theoretical results (Zidek et al., 2002). A related argument is that the daily deAR'd residuals are approximately independent from each other based on the AR structure of the detrended residuals, approximately satisfying the independence assumptions for the responses in (4.1)–(4.3). These arguments also apply to the Chicago's ground–level ozone field, as demonstrated by Figure 4.7 which depicts the partial autocorrelation functions (PACFs) of the detrended residuals, indicating an hourly AR(2) process. That implies the possible choices of dimensions, 2,..., or even 6, for the daily response vectors to achieve adequate temporal separation between them to ensure independence while at the same time, allowing pairwise temporal correlations between hours to be estimated. Choice of the appropriate dimensionality of the response vectors is based on how severe the spatial correlation leakage may result. In other words, we use multivariate AR process for the detrended residuals, avoiding different AR process for responses at different monitoring locations.

For simplicity, let $(i{:}j)$–hour represent the sub–data matrix for the hourly responses from hour i to j across the gauged sites, for $i, j \in \{1, \ldots, p\}$. Suppose one were interested in interpolating say hour 11's ozone level at any fixed ungauged site given all the observed responses at the gauged sites. As we discussed above, the possible response vectors could be the (10:11)–hours, ..., or the (6:11)–hours. For each choice of these response vectors, the spatial correlations between all the gauge sites are estimated using the multivariate BSP approach. Figure 4.8 shows these estimated spatial correlations for the (10:11)–hours, ..., (6:11)–hours response vectors and that of the detrended residuals. Notice how the spatial correlation declines as the dimensions of the response vector increases (leakage). This boxplot shows that the (10:11)–hours response vectors have the smallest number of loss of spatial correlations between the gauged sites, so becoming the right choice for a multivariate data model. This is true for other hours as well, strongly

69
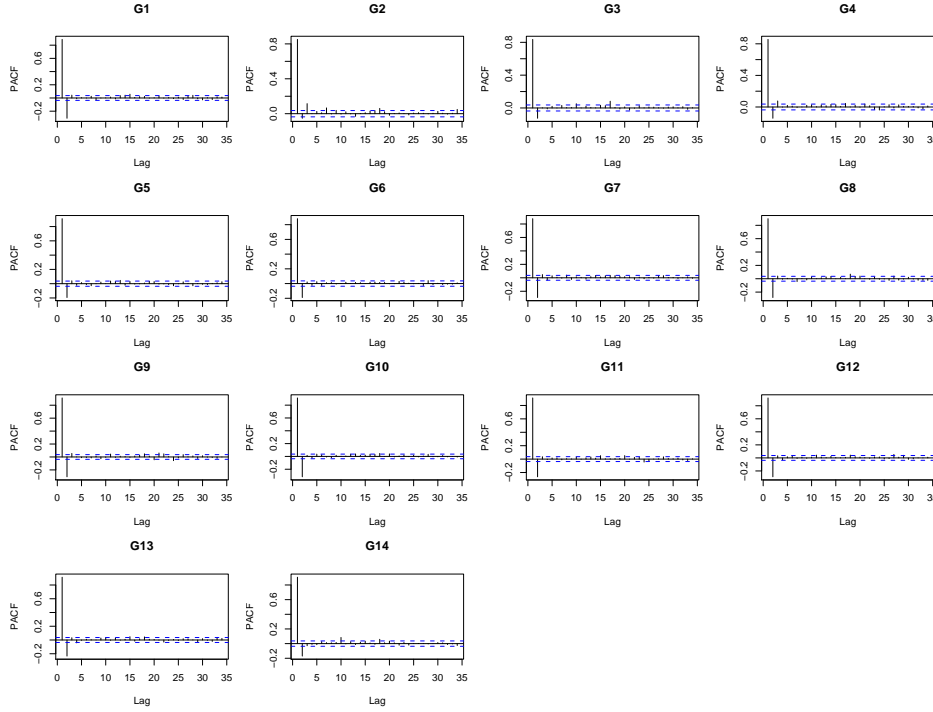
Figure 4.7: The PACF plots for the square–root of hourly ozone concentrations ($\sqrt{ppb}$) at the 14 gauged sites in the Chicago's area AQS database.

supporting the use of 2–hour–block as the response vector. Hence, 2–hour–block data are extracted from the Chicago AQS ozone database (2000) to serve as the multivariate responses in a multivariate BSP model designed to interpolate the hourly ozone concentrations in the field.

Prior to implementing the multivariate BSP approach, a small number of missing measurements are filled in by the spatial regression method. For the model in (4.1)–(4.3), $p = 2$, $n = 123$, $u = 10$, and $g = 14$. The multivariate BSP approach is repeated 24 times by successively cycling the first hour in the two–hour–block through the day to predict the hourly ozone levels at the 10 ungauged sites, and the corresponding 95% pointwise predictive intervals.

For example, suppose the hour of "interest" was hour 11. The response variable at any fixed day, t, and gauged site, j, could be written
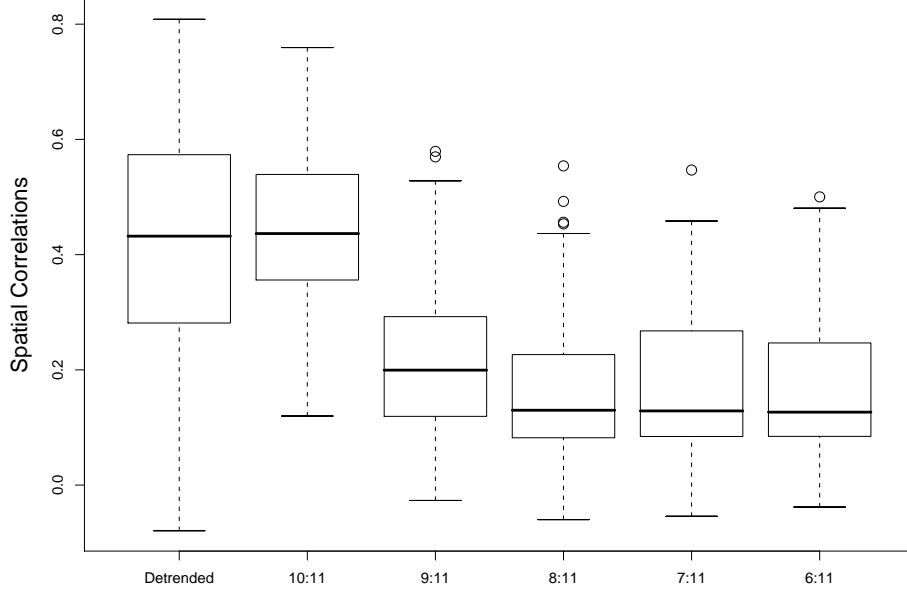
Figure 4.8: Boxplots for the spatial correlations of the detrended residuals (Detrended), and the estimated spatial correlations using the square–root of hourly ozone concentrations during the hours of: 9 A.M. to 10 A.M. (10:11), 8 A.M. to 10 A.M. (9:11), 7 A.M. to 10 A.M. (8:11), 6 A.M. to 10 A.M. (7:11), and 5 A.M. to 10 A.M. (6:11), respectively.

as $\mathbf{Y}_t^{[g^j]} = (\mathbf{Y}_{t,10}^{[g^j]}, \mathbf{Y}_{t,11}^{[g^j]}) : 1 \times p$. Given the BSP model in (4.1)–(4.3) and priors of the parameters, the predictive posterior distribution of $\mathbf{Y}^{[u]}$ is given by (4.7). Notice that $l = 0$ because the response vector contains no missing values. Two covariates, month with four levels and weekday with seven levels, are considered in this approach due to the exploratory data analysis described above for this field, which returns $h = 11$. Let $\mathbf{e}_1 = (0, 1)'$ and $\mathbf{E}_1 = (\mathbf{e}_1', \ldots, \mathbf{e}_1')' : up \times 1$. By a basic property of the multivariate t–distribution and the predictive posterior distribution of $\mathbf{Y}_t^{[u]}$ in (4.17), the posterior distribution of the interested pollutant at ungauged sites is given

by

$$\mathbf{Y}_t^{[u]}\mathbf{E}_1|\mathbf{Y}_t^{(2)}, \mathcal{H} \quad \sim \quad t_{1\times u}(\hat{\mathbf{y}}_t^{[u]}\mathbf{E}_1, \frac{1}{\delta - up - l + 1}P_{t|2}\mathbf{E}_1'\mathbf{\Phi}_{u|g}\mathbf{E}_1,$$
$$\delta - ug - l + 1). \tag{4.20}$$

Let $\mathbf{e}_{0,j}$ be the u–dimensional vector such that the jth entry is 1 and 0 otherwise, for $j = 1, \ldots, u$. Then the unobserved p^{th} pollutant at the ungauged site u_j, $\mathbf{Y}_t^{[u]}\mathbf{E}_1\mathbf{e}_{0,j}$, has the following conditional posterior distribution

$$\mathbf{Y}_t^{[u]}\mathbf{E}_1\mathbf{e}_{0,j}|\mathbf{Y}_t^{(2)}, \mathcal{H} \quad \sim \quad t_{\delta-up-l+1}(\hat{\mathbf{y}}_t^{[u]}\mathbf{E}_1\mathbf{e}_{0,j}, \frac{1}{\delta - up - l + 1}P_{t|2}\mathbf{e}_{0,j}'\mathbf{E}_1'$$
$$\times \mathbf{\Phi}_{u|g}\mathbf{E}_1\mathbf{e}_{0,j}). \tag{4.21}$$

Hence, the predictive posterior mean and variance of the unobserved p^{th} pollutant are given by

$$\mathrm{E}(\mathbf{Y}_t^{[u]}\mathbf{E}_1\mathbf{e}_{0,j}|\mathbf{Y}_t^{(2)}, \mathcal{H}) = \hat{\mathbf{y}}_t^{[u]}\mathbf{E}_1\mathbf{e}_{0,j}$$

and

$$\mathrm{Var}(\mathbf{Y}_t^{[u]}\mathbf{E}_1\mathbf{e}_{0,j}|\mathbf{Y}_t^{(2)}, \mathcal{H}) = \frac{1}{\delta - up - l + 1}P_{t|2}\mathbf{e}_{0,j}'\mathbf{E}_1'\mathbf{\Phi}_{u|g}\mathbf{E}_1\mathbf{e}_{0,j},$$

respectively. Consequently, the pointwise 95% predictive intervals of the unobserved p^{th} pollutant at the ungauged site u_i and fixed day t is given by:

$$\mathrm{E}(\mathbf{Y}_t^{[u]}\mathbf{E}_1\mathbf{e}_{0,j}|\mathbf{Y}_t^{(2)}, \mathcal{H}) \pm t_{\delta-up-l+1}(0.025)(\mathrm{Var}(\mathbf{Y}_t^{[u]}\mathbf{E}_1\mathbf{e}_{0,j}|\mathbf{Y}_t^{(2)}, \mathcal{H}))^{1/2}. \tag{4.22}$$

The above procedure is then repeated 24 times to interpolate the hourly ozone levels in this field. The software used for this multivariate BSP approach is *EnvioStat.1.0*, found at *http://enviro.stat.ubc.ca*.

We next compare our approach with two others, the DLM and NAIVE.

The DLM approach

The dynamic linear model (DLM) is one alternative approach to the spatial interpolation in the Chicago area's hourly ozone field. The DLM and its implementation have been extensively explored in Chapter 2. In the Chicago's AQS ozone database, the total number of time points (i.e., the hours), T, is $123 \times 24 = 2952$. To compare its interpolation results with those of the multivariate BSP, same gauged and ungauged sites are selected as in Section 4.4. Hence the total number of the gauged sites, n, is 14. The initial settings for the DLM are given next, following an investigation of the discount factor for first–order polynomial dynamic models, to optimize this approach.

Our investigation indicates that the appropriate prior for the phase parameters $\mathbf{a} = (a_1, a_2)'$ is $N(\mu_0, \mathbf{\Sigma}_0)$, where $\mu_0 = (1.5, 4.5)'$ and

$$\mathbf{\Sigma}_0 = \begin{pmatrix} 0.0625 & 0 \\ 0 & 0.5625 \end{pmatrix}.$$

The best initial specification for the state parameters turns out to be $\mathbf{m}_0 = (5.15, -0.75\mathbf{1}'_n, 0.05\mathbf{1}'_n)'$ for location while $\mathbf{C}_0$ is a block diagonal matrix with diagonal entries: 0.1304, $0.0158\mathbf{I}_n$, and $0.0003\mathbf{I}_n$.

The first–order polynomial dynamic model, for $t \geq 1$, is then given by:

$$y_t = \beta_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2) \tag{4.23}$$

$$\beta_t = \beta_{t-1} + \omega_t \quad \omega_t \sim N(0, \sigma^2 \sigma_\beta^2), \tag{4.24}$$

and the initial information: $\beta_0 \sim N(0, \sigma_0^2)$. The measurement of the rate of adaptation to new data, i.e., the adaptive coefficient A_t, converges to the constant A as $t \to \infty$, where A is a function of $r = \sigma_\beta^2$, the ratio of the state variance to the system variance ($r \in [0, 1]$) (Harvey, 1984; West & Harrison, 1997). r is also called the signal–to–noise ratio, large values of r implying a clear signal in the system. The adaptive coefficient A reflects how much weight to put on the new data in updating the predictive mean for this simplest DLM: larger value of A represents more weight on the new data; smaller value of A represents less weight on the new data. Consequently, the

discount factor, δ, is defined as $1 - A$. Hence, the larger the discount factor is, the less weight is attached to information provided by the new data in updating the predictive mean at current time; the smaller δ is, the more weight is put on the information of the new data. If $\delta = 0$, the dynamic model would not put any weight on the new data, resulting in a very poor model fit. In other words, there is no signal in the system but only the noise from the measurements. It is impossible to distinguish the signal and the noise in the system. If $\delta = 1$, almost all the weight would be on the new data, resulting in a non–distinguishable signal and noise. West and Harrison (1997) recommend setting the values of δ between 0.8 and 1. A, r, and δ, are related by the equation: $r = A^2(1 - A)^{-1} = (1 - \delta)^2\delta^{-1}$ (West & Harrison, 1997).

δ	0.80	0.85	0.90	0.95	0.99
r	0.05	0.0265	0.0111	0.0026	0.0001

Table 4.1: The relationship between the discount factor (δ) and the signal–to–noise ratio (r).

Table 4.1 shows some of the values of δ and its corresponding r, that is, σ_β^2 in (4.23)–(4.24). This table provides one way to set the initial values of σ_β^2. For the constant DLM model in Chapter 2 given the common variance σ^2 the canonical trend has a Gaussian distribution with variance $\sigma^2\tau_y^2$, the periodic trend with period 24, a Gaussian distribution with variance $\sigma^2\tau_1^2$, and the periodic trend with period 12, a Gaussian distribution with variance $\sigma^2\tau_2^2$. The variability of the canonical trend is assumed to be larger than that of the periodic trends; while the variability of the periodic trend with period 24 is assumed to be larger than that of the periodic trend with period 12. In other words, one assumes that the canonical trend puts more weight on the information provided by the new data than other periodic trends, and the periodic trend with periodic 24 has more weight than the periodic 12 trend. For example, if the discount factors of the canonical, periodic 24, and periodic 12 trends were set to 0.85, 0.90, and 0.95, respectively, the corresponding values of τ_y^2, τ_1^2, and τ_2^2 would be 0.0265, 0.0111, and 0.0026, respectively. Those values, divided by 17, are then selected to be the fixed

values of the model parameters: τ_y^2, τ_1^2, and τ_2^2. [The reason to divide them by 17 were discussed in Section 3.4 to take account of the 17 weeks' time effect.] Both λ_1 and λ_2 are set at 25. The hyperpriors for σ^2 and λ are $IG(2,2)$ and $IG(1,5)$, respectively.

Those initial settings of the DLM in Section 3.2 improve the DLM model's fit in interpolating the responses at ungauged sites. However, it underestimates the variability of the system, leading to a trade–off between the precision and the variability of the spatial interpolator. The interpolation results for Chicago's hourly ozone field are shown in Section 4.4. The software, *GDLM.1.0.*, can be freely downloaded from http://enviro.stat.ubc.ca/dlm, to allow our results to be reproduced or for application in other contexts.

NAIVE approach

One might well suspect that the responses at the nearby gauged sites affect interpolation at ungauged sites. It may seem plausible that for a rough spatio–temporal field, the spatial correlations between the sites are so small that the responses at different sites can be viewed to be approximately independent from each other. This leads to NAIVE approach, i.e., the "interpolated" values at any ungauged site are the observed responses at that gauged site closest to the ungauged one.

Ungauged Site	Gauged Site	GCD (km)
1	1	18.19
2	1	5.74
3	1	6.09
4	4	8.70
5	7	12.99
6	7	19.46
7	7	13.50
8	1	13.97
9	11	15.86
10	10	9.33

Table 4.2: The greatest circle distance (GCD) between the pairs of ungauged sites and their closest gauged site(s) in the Chicago's hourly O_3 field.

Table 4.2 represents the GCD (in km) between the ungauged sites and their closest gauged sites in the Chicago's hourly ozone field. For example, Gauged Site 11 is the closest one to Ungauged Site 9 in terms of the GCD in Table 4.2. The interpolated values at Ungauged Site 9 are the observed responses at Gauged Site 11 in this approach.

NAIVE approach ignores the local characteristics in the field, not taking account of the effects of the covariates, or other factors such as wind and wind directions among the stations. If NAIVE approach had been proved comparable or better than the multivariate BSP and DLM, one would prefer it owing to its great simplicity. Section 4.4 compares NAIVE approach with the multivariate BSP and DLM.

Comparisons and results

This subsection implements the multivariate BSP, DLM, and NAIVE approaches, to interpolate the hourly ozone concentrations at ungauged sites in the Chicago area. These interpolation results are compared with each other to assess the model performance of the multivariate BSP, DLM, and NAIVE approaches. Figures 4.9–4.16 plot the interpolation results of the three approaches at one of the ungauged sites, Ungauged Site 7, during the first week to the sixteenth week. In these four figures, the dots represent the observed values at this ungauged site, the solid and dotdashed lines represent the predictive mean and 95% pointwise predictive intervals for the multivariate BSP approach, the dashed and dotted lines represent the predictive mean and 95% empirical predictive intervals for the DLM approach, and the addition signs represent the interpolated value by NAIVE approach, i.e., the observed values at Gauged Site 7 in this case (refer to Table 4.2). Those empirical predictive intervals of the DLM are wiggly and monotonically increasing as time increases, one feature observed in Section 3.4. On the other hand, the 95% pointwise predictive intervals of the multivariate BSP do not have such wiggly behavior and capture both the long– and short–term trends at Ungauged Site 7. The interpolation of NAIVE approach is not as good as those of the DLM and multivariate BSP for most

time points, but it does occasionally do well when the interpolation results for the other two approaches deviate from the observed responses.
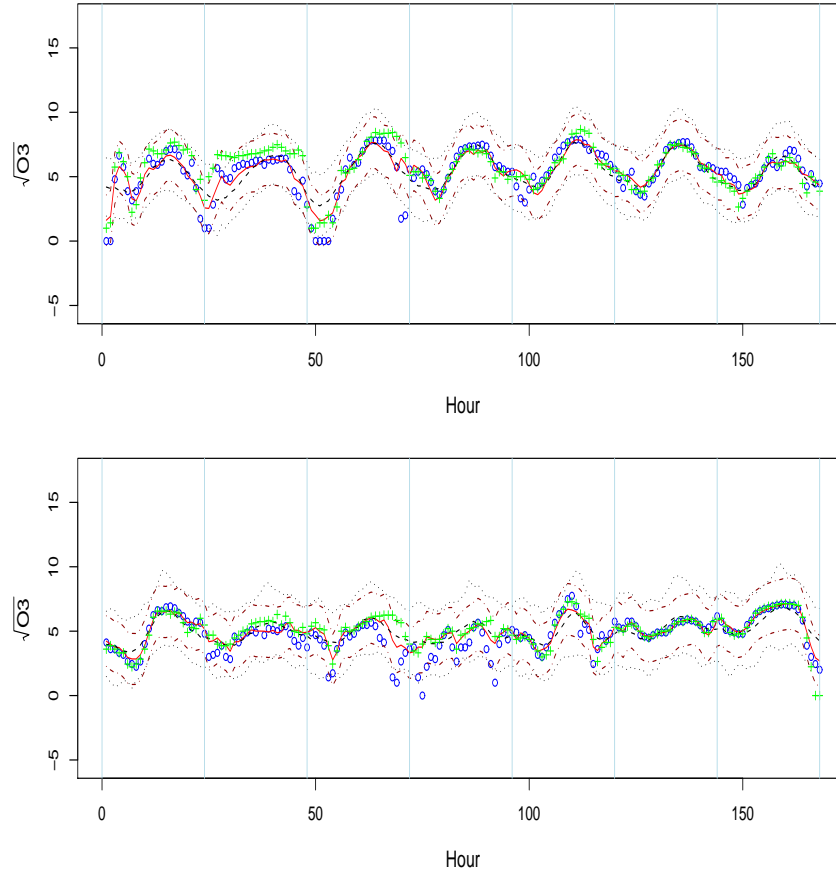
Figure 4.9: Interpolation at Ungauged Site 7 from the 1st week to the 2nd week. The square–root of hourly ozone concentrations are plotted on the vertical axes and hours, on the horizontal axes. [Solid (dot-dashed) lines = interpolation and 95% pointwise predictive intervals by the BSP; dash (dot) lines = interpolation and 95% predictive intervals by the DLM; + = interpolation by NAIVE; and ○ = observations at Ungauged Site 7.]

The closer the ungauged sites are to the gauged ones, the better the expected interpolation performance; that is the assumption upon which these three approaches rest. Figures 4.17–4.18 plot the interpolation results for

77
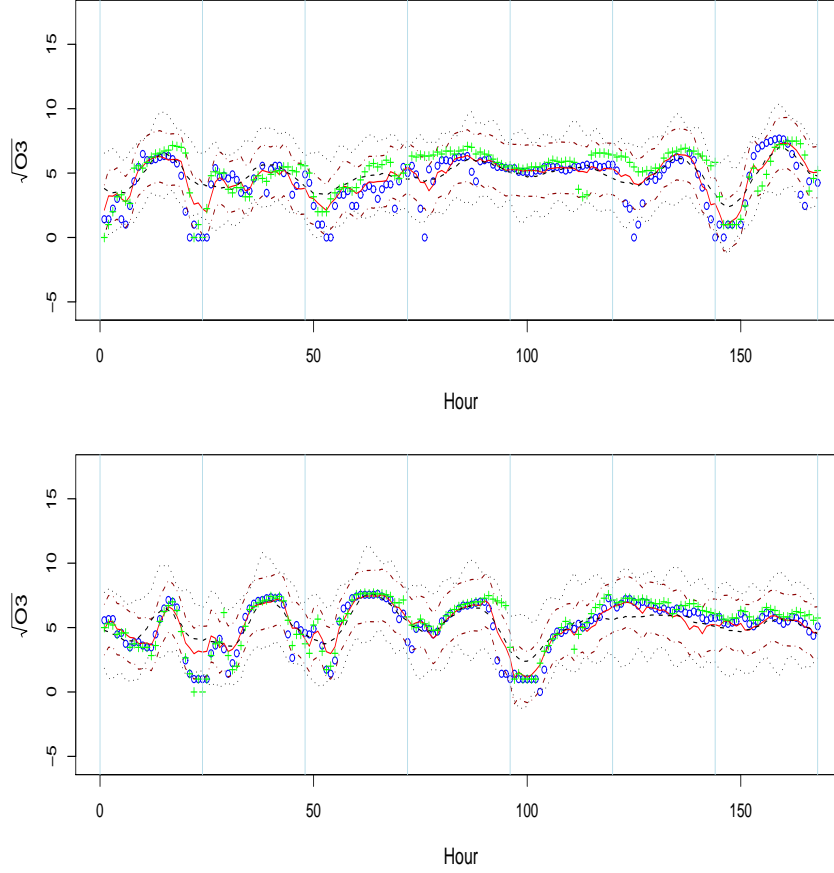
Figure 4.10: Interpolation at Ungauged Site 7 from the 3rd week to the 4th week. The square–root of hourly ozone concentrations are plotted on the vertical axes and hours, on the horizontal axes. [Solid (dot-dashed) lines = interpolation and 95% pointwise predictive intervals by the BSP; dash (dot) lines = interpolation and 95% predictive intervals by the DLM; + = interpolation by NAIVE; and ∘ = observations at Ungauged Site 7.]

the three approaches at Ungauged Site 10 during the 1st and 10th week, respectively. The overestimated predictive variances of the DLM have a monotone increasing trend in the 10th week, compared with those in the 1st week. NAIVE approach performs quite differently from one time to the next. For example, Figure 4.17 shows that NAIVE approach produces good
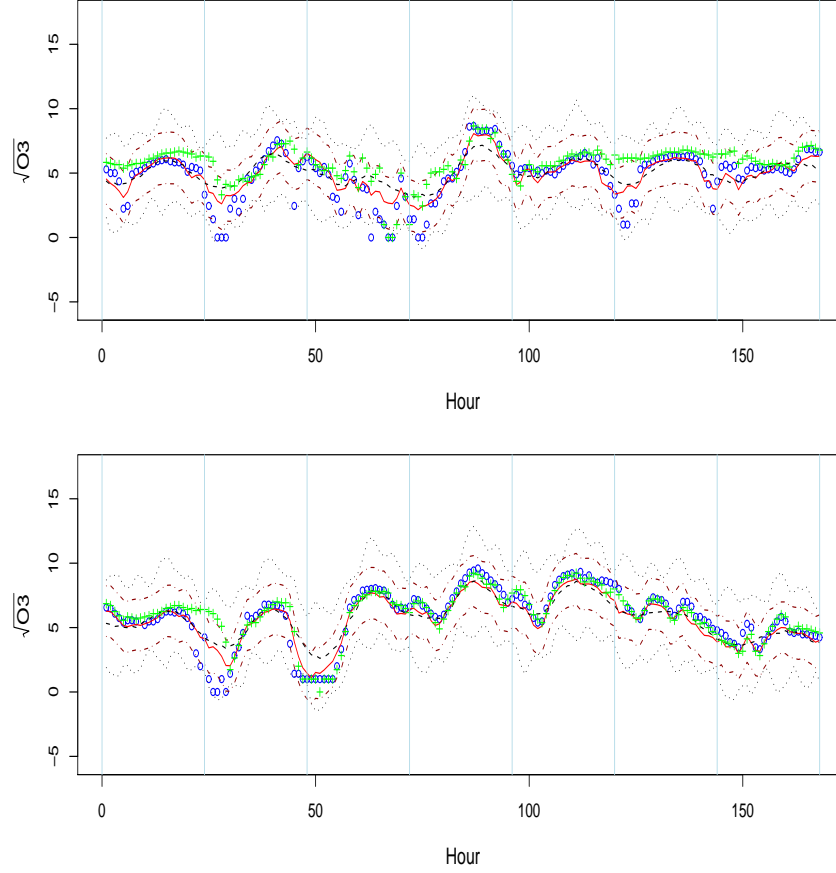
Figure 4.11: Interpolation at Ungauged Site 7 from the 5th week to the 6th week. The square–root of hourly ozone concentrations are plotted on the vertical axes and hours, on the horizontal axes. [Solid (dot-dashed) lines = interpolation and 95% pointwise predictive intervals by the BSP; dash (dot) lines = interpolation and 95% predictive intervals by the DLM; + = interpolation by NAIVE; and ∘ = observations at Ungauged Site 7.]

predictive values that are quite close to the observed responses at Ungauged Site 10 at most times in the 1st week. However, Figure 4.18 shows that the predicted values for NAIVE approach tend to underestimate the observed responses. One therefore suspects that other factors contribute to the spatial correlation between these sites, for instance, covariate effects, such as wind
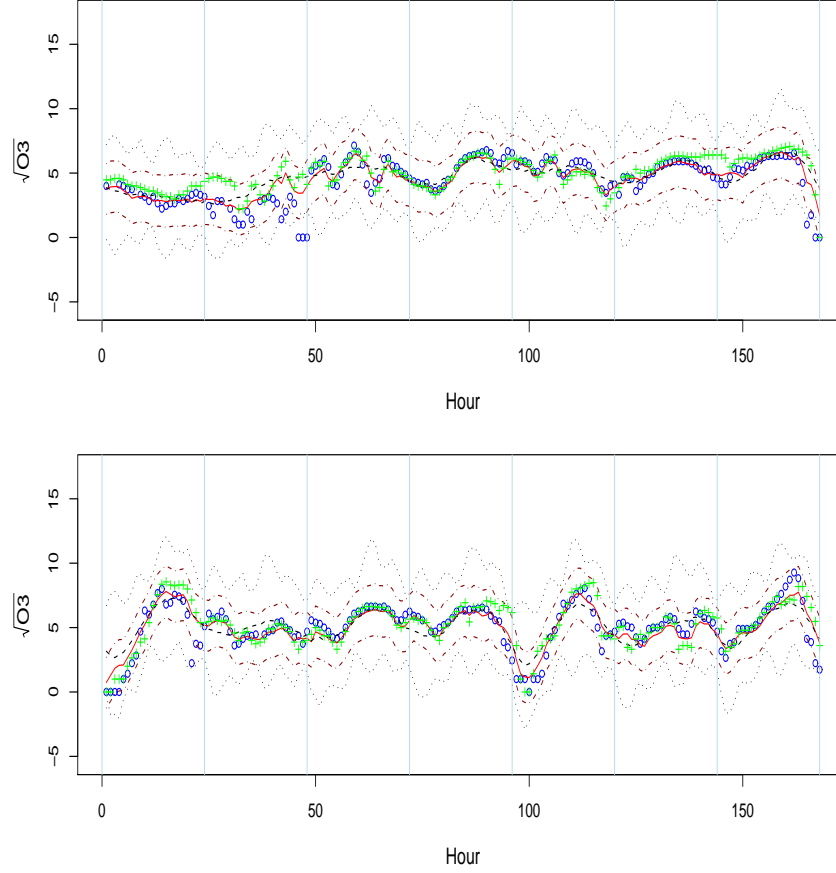
Figure 4.12: Interpolation at Ungauged Site 7 from the 7th week to the 8th week. The square–root of hourly ozone concentrations are plotted on the vertical axes and hours, on the horizontal axes. [Solid (dot-dashed) lines = interpolation and 95% pointwise predictive intervals by the BSP; dash (dot) lines = interpolation and 95% predictive intervals by the DLM; + = interpolation by NAIVE; and ∘ = observations at Ungauged Site 7.]

and its direction. Two close–by sites may be correlated because they lie in the same wind direction; on the other hand, two close–by sites may look independent of one another if their locations lie on a line orthogonal to the wind direction. Other interpolation results, demonstrated in Figures 4.19–4.21, are also examined to check for the association between the response
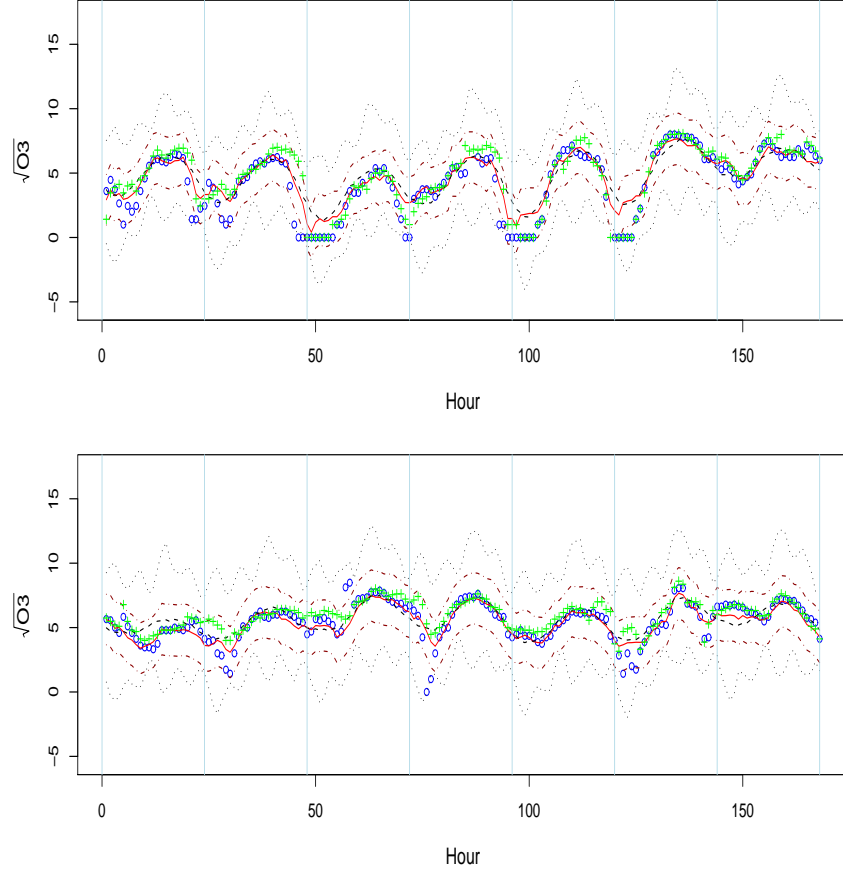
Figure 4.13: Interpolation at Ungauged Site 7 from the 9th week to the 10th week. The square–root of hourly ozone concentrations are plotted on the vertical axes and hours, on the horizontal axes. [Solid (dotdashed) lines = interpolation and 95% pointwise predictive intervals by the BSP; dash (dot) lines = interpolation and 95% predictive intervals by the DLM; + = interpolation by NAIVE; and ∘ = observations at Ungauged Site 7.]

variables at the ungauged site and its closest gauged neighbour.

The marginal posterior distribution of the pth pollutant tends to overestimate the predictive variance of the process, as observed in Figures 4.17–4.21. The posterior ellipsoid credible regions, constructed from the joint posterior distributions of the p pollutants, provide well calibrated predictive results
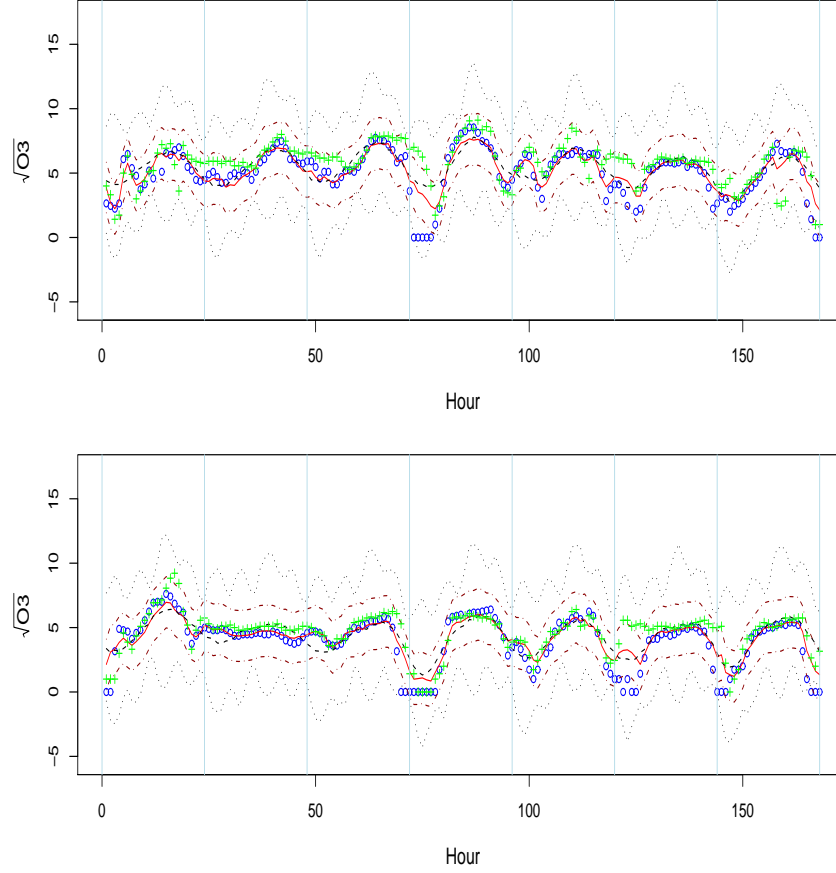
Figure 4.14: Interpolation at Ungauged Site 7 from the 11th week to the 12th week. The square–root of hourly ozone concentrations are plotted on the vertical axes and hours, on the horizontal axes. [Solid (dotdashed) lines = interpolation and 95% pointwise predictive intervals by the BSP; dash (dot) lines = interpolation and 95% predictive intervals by the DLM; + = interpolation by NAIVE; and ○ = observations at Ungauged Site 7.]

using the multivariate BSP approach. Figure 4.22 plots the empirical ellipsoid coverage probabilities across the ungauged sites at various nominal levels. Their averages across the ungauged sites are listed in Table 4.3, showing slightly underestimated predictive variances.

Table 4.4 shows the mean square predictive error (MSPE) at each un-
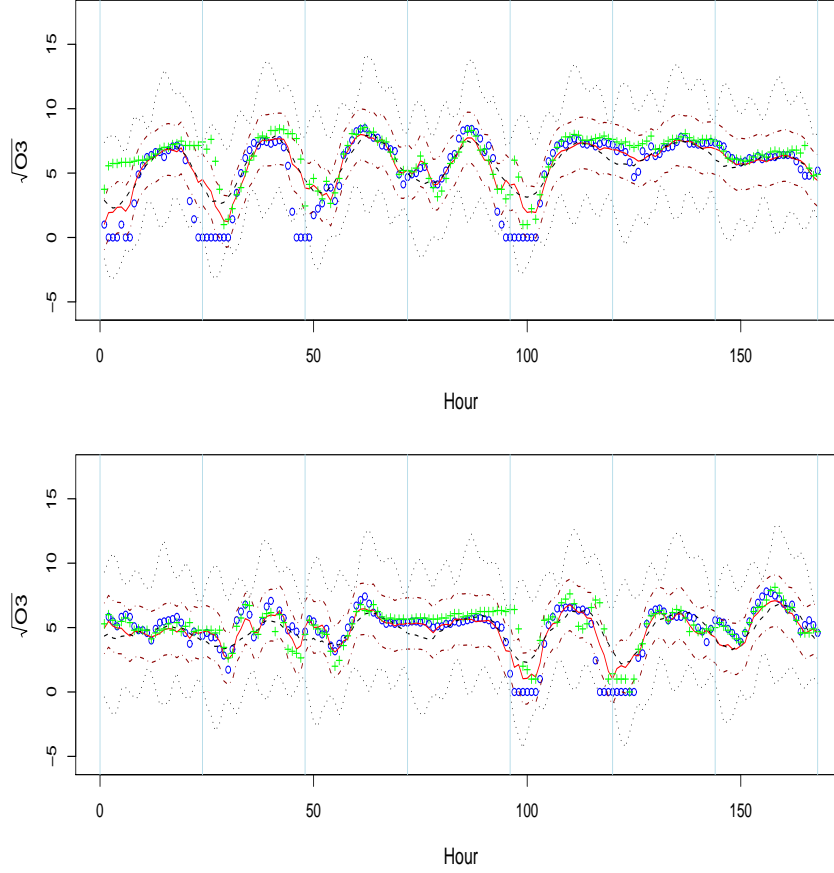
Figure 4.15: Interpolation at Ungauged Site 7 from the 13th week to the 14th week. The square–root of hourly ozone concentrations are plotted on the vertical axes and hours, on the horizontal axes. [Solid (dotdashed) lines = interpolation and 95% pointwise predictive intervals by the BSP; dash (dot) lines = interpolation and 95% predictive intervals by the DLM; + = interpolation by NAIVE; and ○ = observations at Ungauged Site 7.]

gauged site obtained by the three approaches. For all the ungauged sites, the NAIVE approach has the largest MSPE and has the poorest model performance of the three. Except for Ungauged Sites 2 and 8, the multivariate BSP has the smallest values of the MSPE among these three approaches, overall the best model performance among the three. At Ungauged Sites
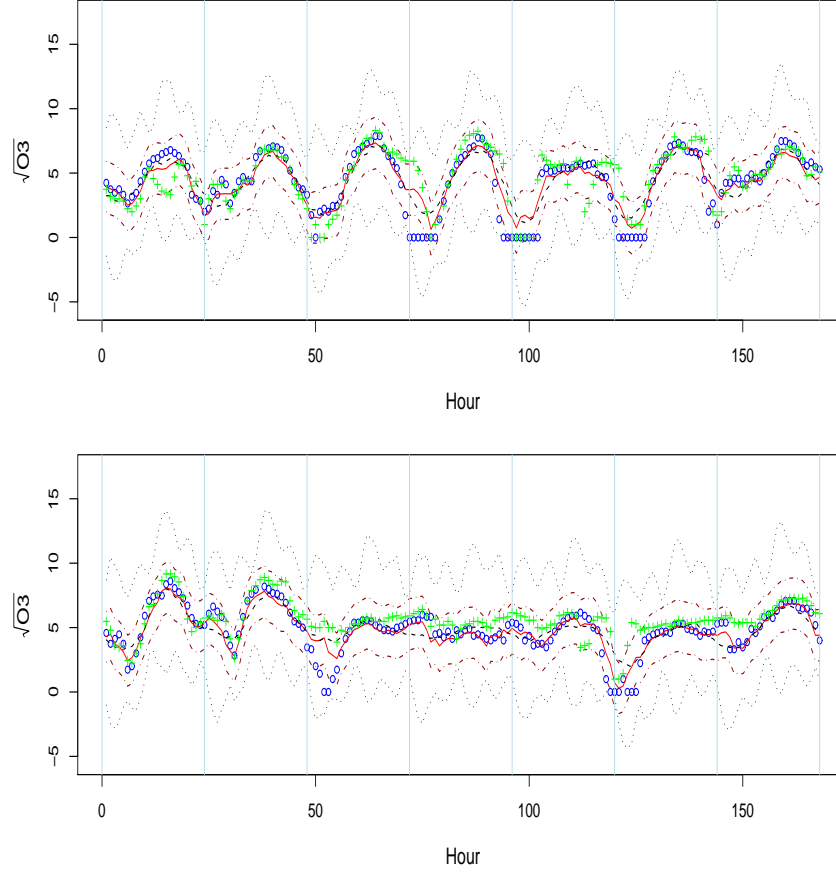
Figure 4.16: Interpolation at Ungauged Site 7 from the 15th week to the 16th week. The square–root of hourly ozone concentrations are plotted on the vertical axes and hours, on the horizontal axes. [Solid (dotdashed) lines = interpolation and 95% pointwise predictive intervals by the BSP; dash (dot) lines = interpolation and 95% predictive intervals by the DLM; + = interpolation by NAIVE; and ∘ = observations at Ungauged Site 7.]

2 and 8, the MSPE of the DLM is slightly smaller than that of the multivariate BSP, implying possible local variations in the characteristics of the ungauged sites. In other words, one might well add a nugget effect in the multivariate BSP approach to account for the variations among the response variables.
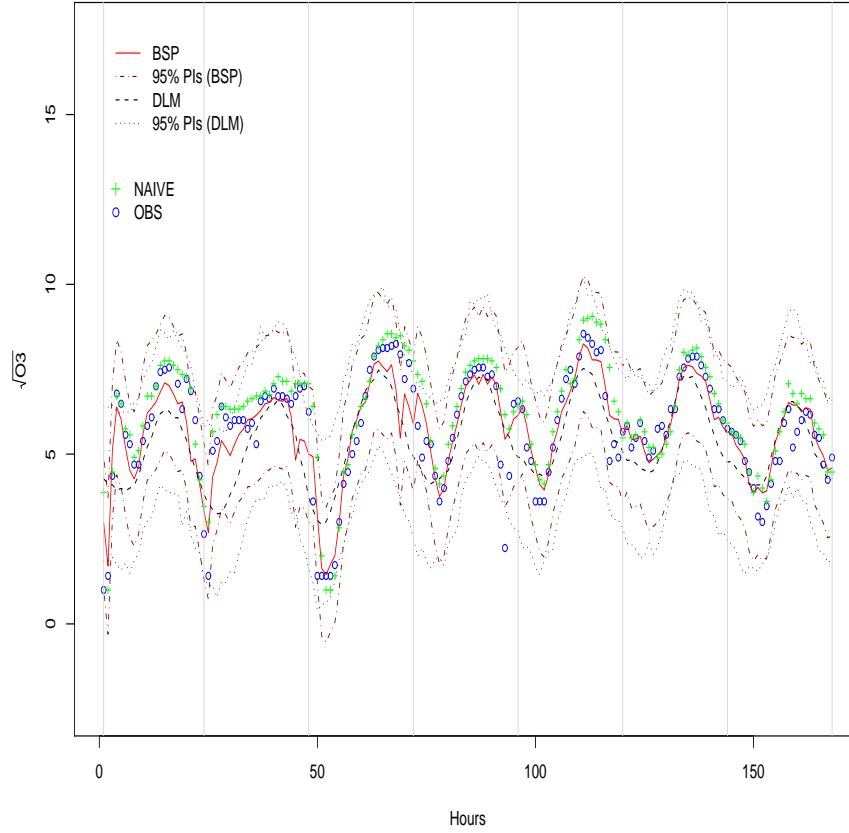
Figure 4.17: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) during the 1$^{\text{st}}$ week, the interpolation using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Ungauged Site 10.

Figures 4.23 and 4.24 plot the ratios of the MSPEs of NAIVE and the DLM to that of the multivariate BSP, respectively. Figure 4.23 shows that the interpolation for the multivariate BSP has smaller MSPE than that for NAIVE approach over all seventeen weeks and ten ungauged sites. Figure 4.24 plots the ratio of the MSPE of the DLM interpolator to that of the multivariate BSP, showing that better model performance of the latter over all 17 weeks and most ungauged sites.
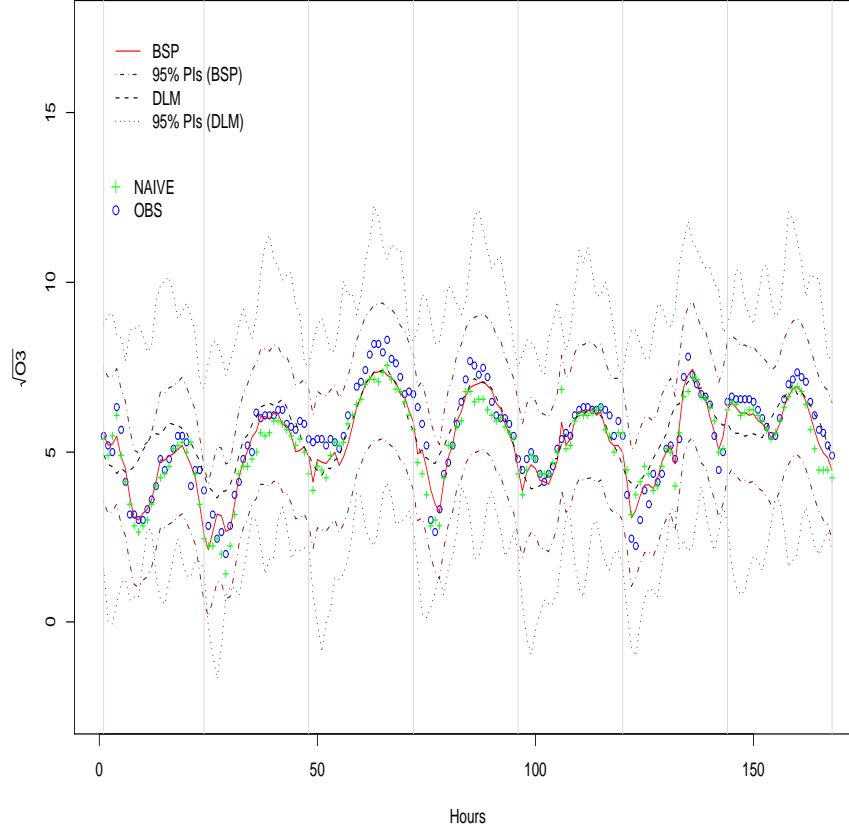
Figure 4.18: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) during the 10th week, the interpolation using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Ungauged Site 10.

The coverage probabilities of the multivariate BSP and DLM are computed at the 95% nominal level as a further assessment of model performance. Figures 4.25 and 4.26 show the side–by–side boxplots of these coverage probabilities over 10 ungauged sites and 17 weeks, respectively. The coverage probabilities of the DLM are much higher than the 95% nominal level, for most weeks and ungauged sites. In fact, the pointwise coverage probabilities of the multivariate BSP tend to be closer to that nominal level

86
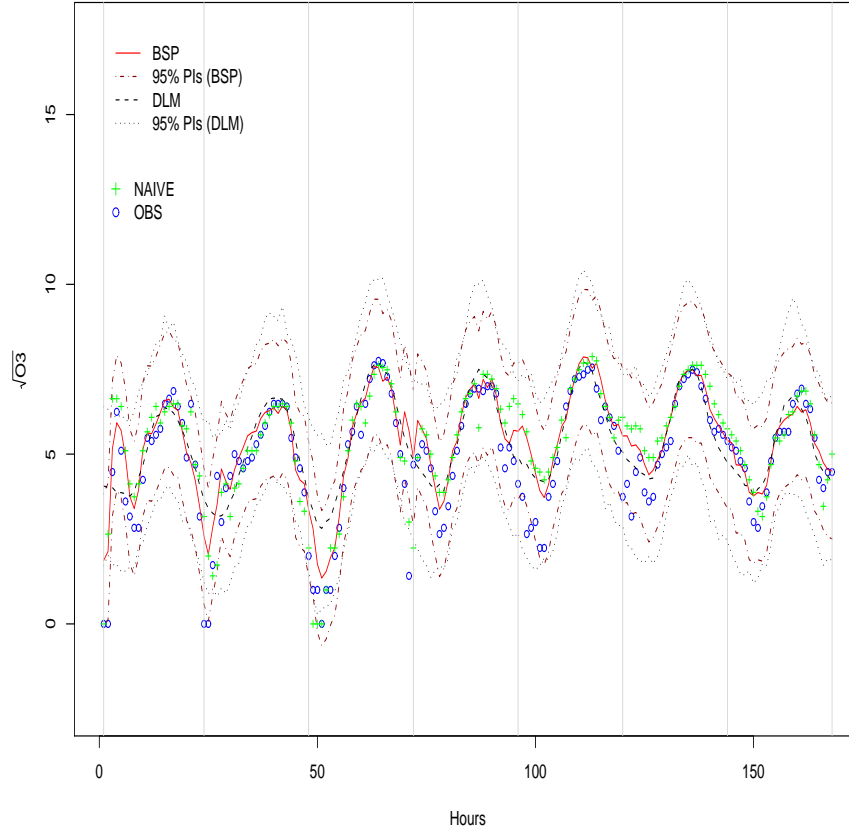
Figure 4.19: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) during the 1st week, the interpolation using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Ungauged Site 9.

than those of the DLM.

Comparing both the MSPEs and coverage probabilities from the three approaches, NAIVE approach has the worst model performance among these three; the DLM approach does better than NAIVE, but not as good as the multivariate BSP. Overall, the multivariate BSP has the best model performance among these three.

Temporal prediction is addressed in the next section by using the mul-
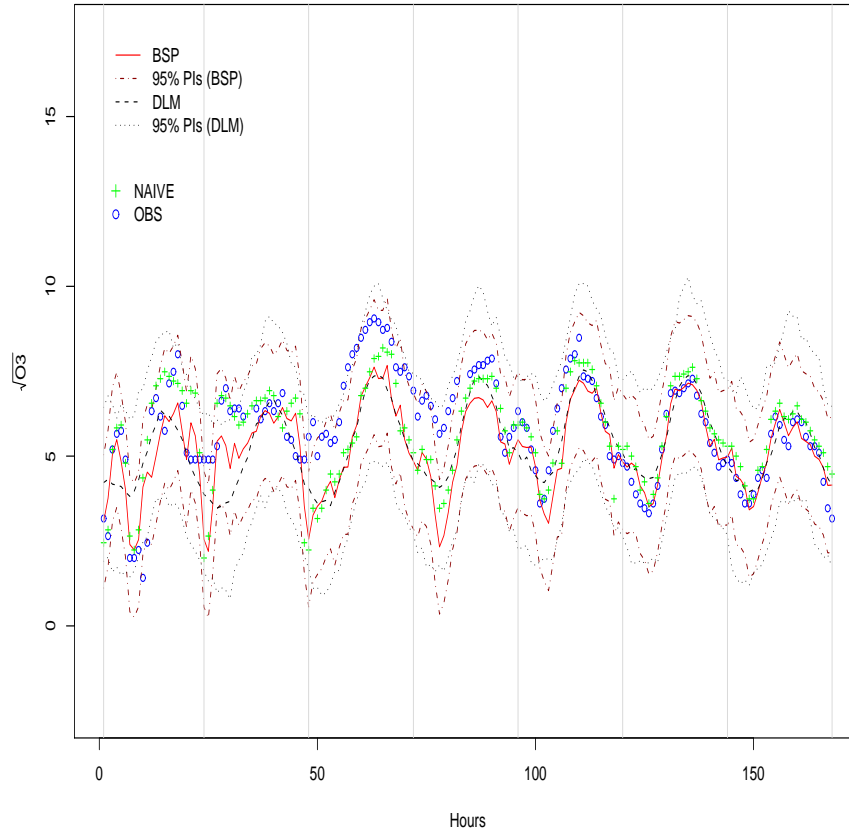
Figure 4.20: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) during the 1$^{\text{st}}$ week, the interpolation using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Ungauged Site 8.

tivariate BSP approach in the Chicago area's hourly ozone field. Moreover, that model is assessed by comparing with the other two approaches: the DLM and NAIVE* (see Section 5.3).
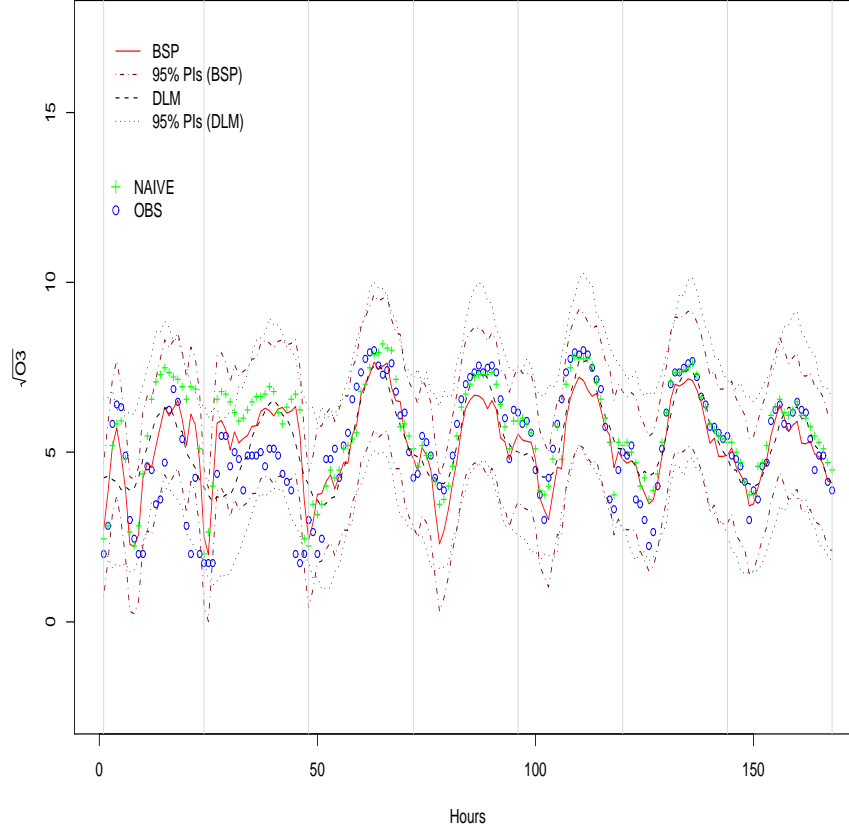
Figure 4.21: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) during the 1$^{\text{st}}$ week, the interpolation using the multivariate BSP, DLM and NAIVE approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Ungauged Site 2.

4.5   Spatial Leakage in the DLM

Zidek et al. (2002) address the crucial problem caused by the spatial leakage, or the space–time interaction problem in spatially interpolating the hourly PM_{10} in Vancouver area using the univariate approach. The spatial leakage occurred when substantial amount of between–site cross–correlations exists until certain time lag. The whitened residuals used in their model come from two steps: (i) firstly obtaining the detrended residuals by taking off
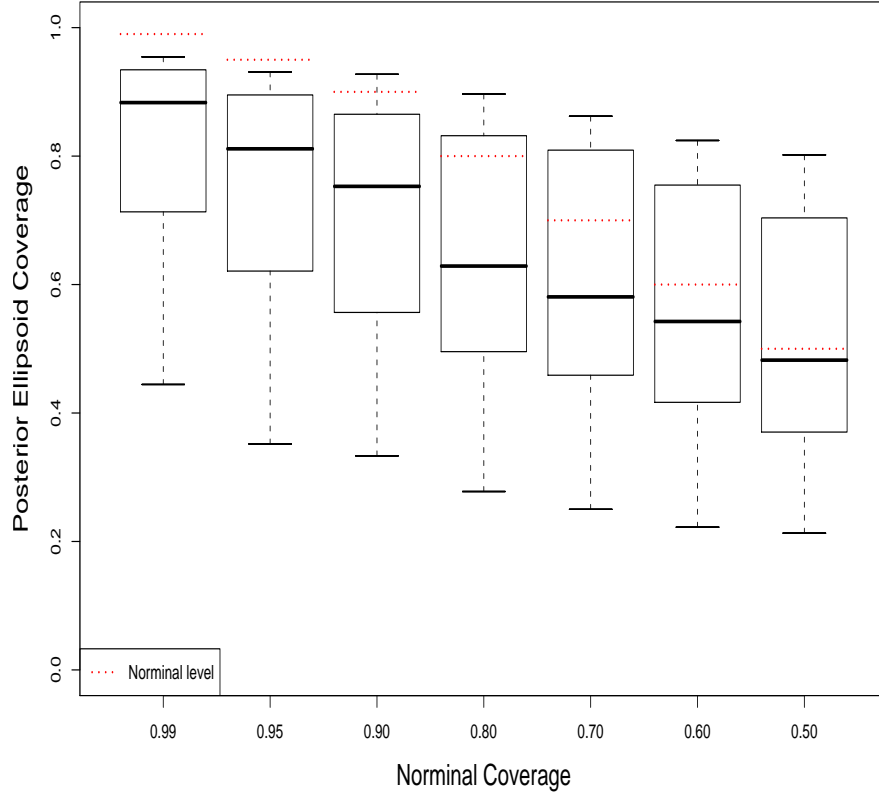
Figure 4.22: Boxplot of the simultaneous posterior ellipsoid credibility regions at various nominal levels.

the covariates effect from the response variables, for fixed site and varying time points; (ii) then the deAR'd (or whitened) residuals can be obtained after taking account the temporal dependence into the detrended residuals.

To be more specific, suppose $Y(\mathbf{s}, t)$ represents the response variable at site $\mathbf{s}$ and time t, for $\mathbf{s} \in \{\mathbf{s}_1, \dots, \mathbf{s}_p\}$ and $t = 1, \dots, n$. Let $\mathbf{z}_t : 1 \times l$ be an l-dimensional covariates vector at time t for all sites and $\beta : 1 \times l$, the regression coefficients corresponding to the l covariates. The detrended

Nominal Coverage (%)	99	95	90	80	70	60	50
Ellipsoid Credible Coverage (%)	82	75	71	65	62	57	53

Table 4.3: The posterior ellipsoid coverage probabilities at various nominal levels.

Ungauged Site	MSPE(NAIVE)	MSPE(DLM)	MSPE(BSP)
1	2.96	1.79	1.16
2	2.17	1.55	1.61
3	1.62	1.80	1.30
4	1.65	1.46	0.94
5	2.52	1.92	1.03
6	3.08	1.85	0.99
7	2.13	1.60	0.97
8	3.74	2.62	2.67
9	1.42	1.63	1.01
10	0.46	0.87	0.38

Table 4.4: The mean square predictive error (MSPE) at ungauged sites of the multivariate BSP, DLM, and NAIVE approaches.

residuals, $E(\mathbf{s}, t)$, are given by

$$E(\mathbf{s}, t) = Y(\mathbf{s}, t) - \mathbf{z}_t \beta. \qquad (4.25)$$

Assuming an AR(q) process for the detrended residuals, the deAR'd residuals are denoted by $e(\mathbf{s}, t)$, and given by

$$E(\mathbf{s}, t) = \sum_{i=1}^{q} \alpha_i E(\mathbf{s}, t - i) + e(\mathbf{s}, t). \qquad (4.26)$$

Theoretical results in Zidek et al. (2002) show the existence of potential spatial correlations leakage for an AR(1) (detrended) process. The connections between Le–Zidek approach and the DLM allow one use the former modelling in the DLM framework. Moreover, as a specific case of state–space model, the DLM framework can be treated as a bridge to unify the Le–Zidek approach and the general state–space modelling. Our objective is to investigate potential spatial correlations leakage in the general state–
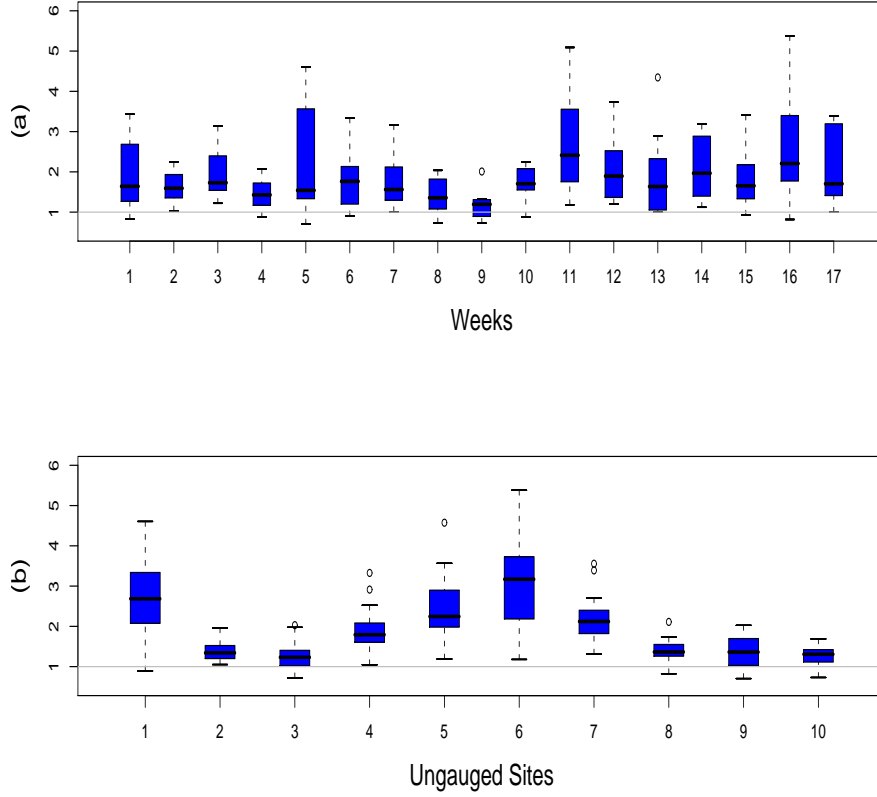
Figure 4.23: The ratio of MSPE of the interpolation by NAIVE to that of the multivariate BSP for each of: (a) the 17 weeks; and (b) the 10 ungauged sites.

space modelling for this approach. The following content addresses the work for an AR(p) process in (4.26). To reformulate the Le–Zidek approach as the DLM, we can assume $\mathbf{z}_t$ itself having a random Gaussian process with mean $\mathbf{0}$ and very small variance matrix, $\sigma_z^2 \mathbf{I}_l$. Of course, we can also assume that $\mathbf{z}_t$ is fixed at each time point t, exactly the same as in Zidek et al. (2002). In other words, we consider two cases here:

Case (i) $\mathbf{z}_t$ itself has a stochastic process, for $t = 1, \ldots, n$, that

$$\mathbf{z}_t = \mathbf{z}_{t-1} + \mathbf{w}_t^z \quad \mathbf{w}_t^z \sim N(\mathbf{0}, \sigma_z^2 \mathbf{I}_l). \tag{4.27}$$
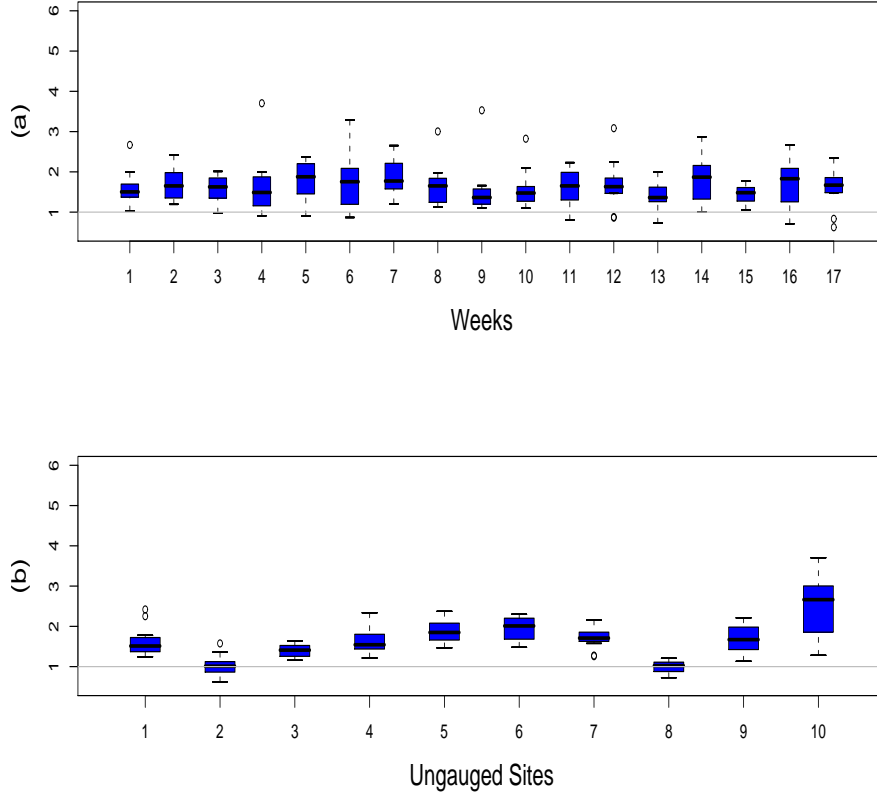
Figure 4.24: The ratio of MSPE of the interpolation by the DLM to that of the multivariate BSP for each of: (a) the 17 weeks; and (b) the 10 ungauged sites.

Case (ii) $\mathbf{z}_t$ is fixed at fixed time point t, for $t = 1, \ldots, n$.

For **Case (i)**, models (4.25)–(4.27) can be normalized in the DLM as follows:

$$Y(\mathbf{s}, t) = \mathbf{F}'\theta(\mathbf{s}, t) \tag{4.28}$$

$$\theta(\mathbf{s}, t) = \mathbf{G}\theta(\mathbf{s}, t - 1) + \mathbf{w}_t, \tag{4.29}$$

where $\mathbf{F}' : 1 \times (q+1) = (1, 1, 0, \ldots, 0), \theta(\mathbf{s}, t)' : 1 \times (q+1) = (\mathbf{z}_t\beta, E(\mathbf{s}, t), E(\mathbf{s}, t-$
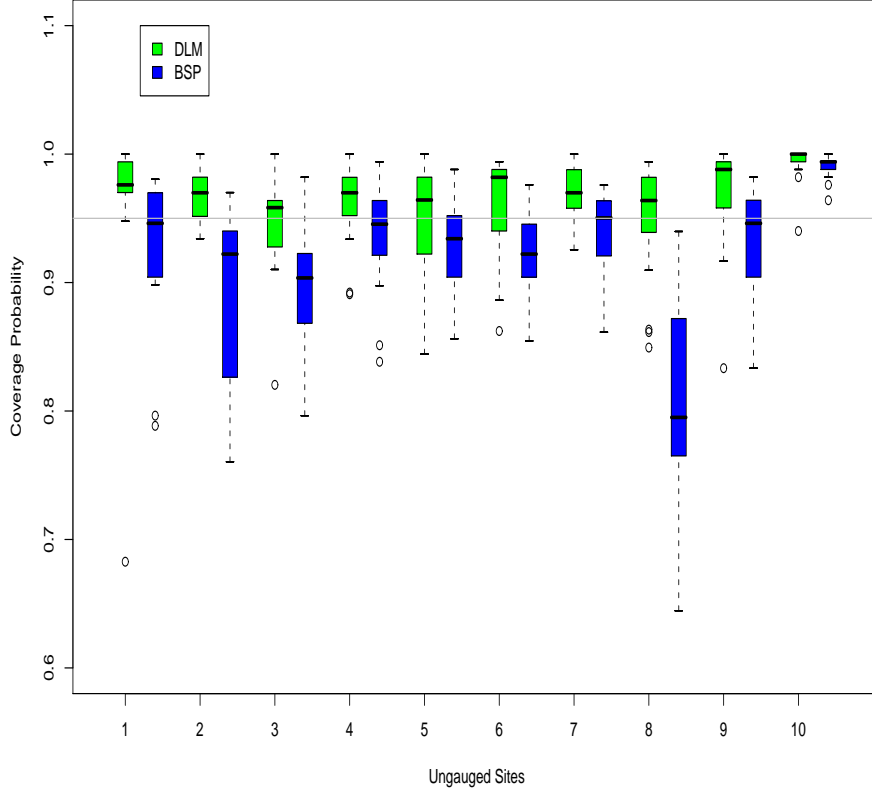
Figure 4.25: Side–by–side boxplots of the coverage probabilities of the multivariate BSP and DLM approaches plotted against the 10 ungauged sites, respectively.

$1), \ldots, E(\mathbf{s}, t - q + 1)), \mathbf{w}'_t : 1 \times (q + 1) = (\mathbf{w}^z_t \beta, e(\mathbf{s}, t), 0, \ldots, 0)$ and

$$\mathbf{G} : (q+1) \times (q+1) = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & \alpha_1 & \alpha_2 & \ldots & \alpha_q \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}.$$

For **Case (ii)**, we can also formalize models (4.25)–(4.27) same as above
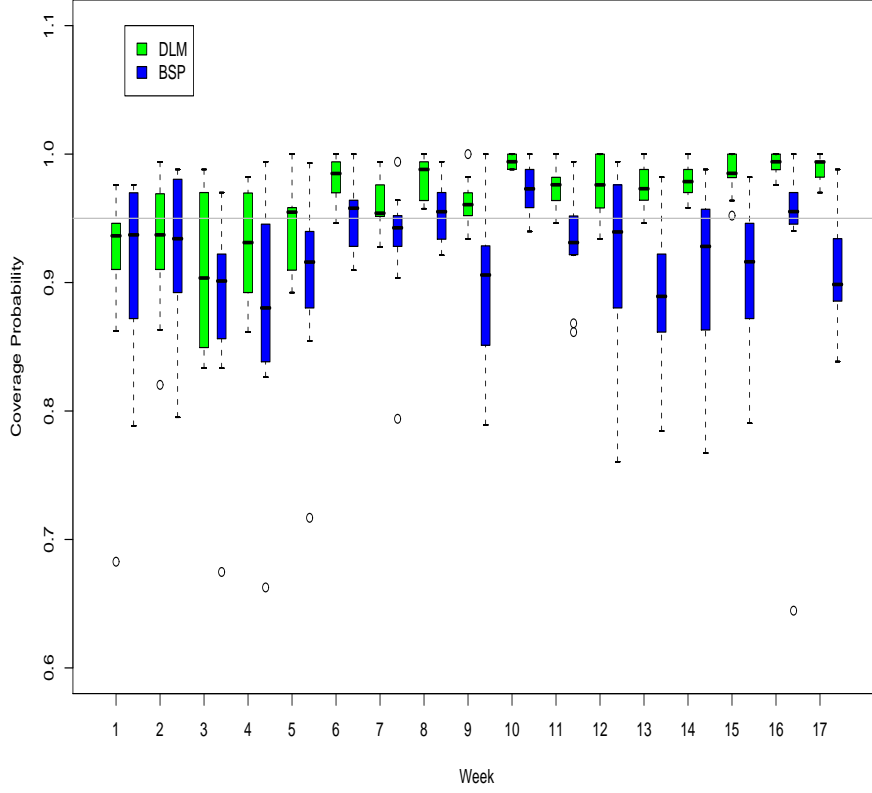
Figure 4.26: Side–by–side boxplots of the coverage probabilities of the multivariate BSP and DLM approaches plotted against the time span of 17 weeks, respectively.

except the first entry in $\mathbf{F}'$ to be $\mathbf{z}_t((\mathbf{z}_t^*)'\mathbf{z}_t^*)^{-1}(\mathbf{z}_t^*)'$ and the $(1,1)$ entry in $\mathbf{G}$, $\mathbf{z}_t^*((\mathbf{z}_{t-1}^*)'\mathbf{z}_{t-1}^*)^{-1}\mathbf{z}_{t-1}^*$, and $\mathbf{z}_t^* = \mathbf{z}_t - \sum_{i=1}^q \alpha_i\mathbf{z}_{t-i}$.

For both cases, let $\mathbf{c}_i' : 1 \times (q+1)$ such that its ith entry being 1 and others, 0, for $i = 1, \ldots, q+1$. Then we have

$$
\begin{aligned}
E(\mathbf{s}, t-i) &= \mathbf{c}_{i+2}'\theta(\mathbf{s}, t) & (4.30)\\
e(\mathbf{s}, t) &= \mathbf{c}_2'\mathbf{w}(\mathbf{s}, t), & (4.31)
\end{aligned}
$$

for $i = 0, \ldots, q - 1$.

Zidek et al. (2002) shows that the spatial correlation leakage occurs for the AR(1) process if $\mathrm{Cor}(E(\mathbf{s}, t-1), e(\mathbf{s}', t)) \neq 0$. Hence the lag–1 between–site cross–covariance for any two sites $\mathbf{s}$ and $\mathbf{s}'$ can be written as:

$$
\begin{aligned}
\mathrm{Cov}(E(\mathbf{s}, t-1), e(\mathbf{s}', t)) &= \mathrm{Cov}\{\mathbf{c}_3' \theta(\mathbf{s}, t), \mathbf{c}_2' \mathbf{w}(\mathbf{s}', t)\} \\
&= \mathbf{c}_3' \mathrm{Cov}\{\theta(\mathbf{s}, t), \mathbf{w}(\mathbf{s}', t)\} \mathbf{c}_2 \\
&= \mathbf{c}_3' \mathrm{Cov}\{\mathbf{G}\theta(\mathbf{s}, t-1) + \mathbf{w}(\mathbf{s}, t), \mathbf{w}(\mathbf{s}', t)\} \mathbf{c}_2 \\
&= \mathbf{c}_3' \mathrm{Cov}(\mathbf{w}(\mathbf{s}, t), \mathbf{w}(\mathbf{s}', t)) \mathbf{c}_2 \\
&= \mathbf{c}_3' \mathbf{W}_t(\mathbf{s}, \mathbf{s}') \mathbf{c}_2, \quad\quad\quad (4.32)
\end{aligned}
$$

where $\mathbf{W}_t : p(p+1) \times p(p+1)$ is the covariance matrix for the vector $\mathbf{w}_t : p(p+1) \times 1 = (\mathbf{w}(\mathbf{s}_1, t)', \ldots, \mathbf{w}(\mathbf{s}_p, t)')'$ at fixed time point t and $\mathbf{W}_t(\mathbf{s}, \mathbf{s}') : (p+1) \times (p+1)$, the covariance matrix for $\mathbf{w}(\mathbf{s}, t)$ and $\mathbf{w}(\mathbf{s}', t)$ for fixed t. Therefore, the lag–1 between–site cross–correlation is not zero if $(\mathbf{W}_t(\mathbf{s}, \mathbf{s}'))_{32} \neq 0$.

In general, for $i = 1, \ldots, q-1$, consider the covariance between $E(\mathbf{s}, t-i)$ and $e(\mathbf{s}', t)$, we then have

$$
\begin{aligned}
\mathrm{Cov}(E(\mathbf{s}, t-i), e(\mathbf{s}', t)) &= \mathrm{Cov}(\mathbf{c}_{i+2}' \theta(\mathbf{s}, t), \mathbf{c}_2' \mathbf{w}(\mathbf{s}', t)) \\
&= \mathbf{c}_{i+2}' \mathrm{Cov}(\theta(\mathbf{s}, t), \mathbf{w}(\mathbf{s}', t)) \mathbf{c}_2 \\
&= \mathbf{c}_{i+2}' \mathbf{W}_t(\mathbf{s}, \mathbf{s}') \mathbf{c}_2. \quad\quad\quad (4.33)
\end{aligned}
$$

Hence, the lag–i between–site cross–correlation is not zero if $(\mathbf{W}_t(\mathbf{s}, \mathbf{s}'))_{i+2,2} \neq 0$. This result implies that the spatial correlation leakage occurs if, for $i = 1, \ldots, q-1$, $\mathbf{W}_t(\mathbf{s}, \mathbf{s}')_{i+2,2} \neq 0$ for any two sites $\mathbf{s}$ and $\mathbf{s}'$. It means that the DLM framework cannot avoid the spatial correlation leakage problem for the AR(q) process if such conditions are satisfied.

4.6 Conclusion and Discussion

The multivariate BSP approach can be used to overcome the spatial correlation leakage problem for the filtered residuals between the sites; that other–

wise tends to incorporate a significant temporal structure in those residuals.

Unlike the DLM, the multivariate BSP approach does not assume stationarity and isotropy for the underlying process. The multivariate BSP can incorporate the uncertainty about hyperparameters through fully hierarchical Bayesian modelling. On the other hand, the DLM approach needs some hyperparameters to be fixed to achieve similar objective. The multivariate BSP can provide ellipsoid credible regions for multiple pollutants and simultaneous prediction; the DLM provides predictive intervals for single pollutant only. The multivariate BSP is much more computationally efficient than the DLM. Unlike the latter it can interpolate hourly ground–level ozone over a large number of monitoring stations, an impossible task for the DLM. In other words, it is computationally scalable.

Chapter 5

Multivariate Bayesian Spatial Prediction and Its Temporal Prediction

Not only are people interested in interpolating responses at ungauged sites given all the information at gauged sites, but also in temporally predicting them at these gauged sites. In fact, 24–hour ahead ozone forecasting is now commonly done in many urban areas. The modeller may be asked: "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?"

The multivariate BSP approach can be adapted to answer such questions by taking a 24–dimensional multivariate response variable formed from the daily 24 hourly univariate response variables, treated like 24 "species" or "pollutants"; each entry therein represents one measurement for each of the successive 24 hours. The multivariate model is needed due to the fact that the daily dependence structure of the 24–dimensional multivariate response variable invalidates the row independence assumption of the deAR'd residuals in the multivariate BSP approach.

To apply the BSP, one must obtain estimates of the hyperscale spatial covariance of the 24 pollutants given all the observed response variables. However, the sequence of the 24–dimensional vectors of hourly responses does not meet the BSP's assumption of independence. To overcome this difficulty, subsequences of vectors are systematically extracted to contain responses separated by 24 hours. This subset of data is then used to predict

each one of the 24 response variables at the future day.

More specifically, two sub–data matrices are formed by selecting these 24–dimensional vectors, one sequence for odd days and a second for even days. Model assumptions hold for each sequence, allowing hyperparameters at gauged sites to be estimated by the EM–algorithm via the *EnviRo.stat* software cited earlier. Each "pair" of estimated hyperparameter sets can then be averaged to form approximate "estimates" of hyperparameters given all the data. Finally, the one–step ahead prediction is implemented at gauged sites given those estimates of hyperparameters and observed responses. However, to get the 24–hour ahead forecast, obstacles remain that we now turn to.

In fact, Section 5.1 demonstrates how to construct the above sequences and the corresponding subdata matrices along with the multivariate settings of the BSP model to predict each one 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 5.2 illustrates the *h*–step ahead prediction by the DLM approach. Section 5.3 presents the one–day ahead temporal prediction by NAIVE* approach. Section 5.4 shows the results and comparisons of the one–day ahead prediction by the three approaches at gauged sites.

5.1   The Multivariate BSP Approach

Suppose $Y_{t,i}^{[g_j^m]}$ represents the unobserved i^{th} response variable at day t, gauged site j, and $Y_{t,i}^{[g_j^o]}$, the observed response variable, for $t = 1, \ldots, T$, $i = 1, \ldots, p$, and $j = 1, \ldots, g$. Specifically in the Chicago's AQS database, $T = 120$, $p = 24$, and $g = 14$. Two cases are considered here to predict the response variable at say (i) 11 P.M., or (ii) any single hour during the 0 A.M. to 10 P.M. period on day 121, the last day in the observation sequence, whose data we set aside to be used for assessment.

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

Call the corresponding multivariate BSP model "Model–1". One of the subdata matrices can be formed by $\{\mathbf{U}_t^{(1)}\} : 1 \times gp$, where $\mathbf{U}_t^{(1)} = (Y_{2t-1,1}^{[g_1^o]}, \ldots, Y_{2t-1,p}^{[g_g^o]})$; the other by $\mathbf{V}_t^{(1)} : 1 \times gp$, where $\mathbf{V}_t^{(1)} = (Y_{2t,1}^{[g_1^o]}, \ldots, Y_{2t,p}^{[g_g^o]})$, for $t = 1, \ldots, 59$. In other words, the observed response variables from the 1ˢᵗ day to the 119ᵗʰ 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 121st day, for $k = 2, \ldots, p$.**

  Call the corresponding multivariate BSP model "Model–k". One of the subdata matrices can be formed by $\{\mathbf{U}_t^{(k)}\} : 1 \times gp$, 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 $\{\mathbf{V}_t^{(k)}\}$, 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 weekdays, 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. This weekday effect is removed from the $\{\mathbf{U}_t^{(k)}\}$s and $\{\mathbf{V}_t^{(k)}\}$s, respectively, and the software, *EnviRo.stat,* used to implement these 24 models.

Let $\mathbf{Y} = (\mathbf{Y}^{[u]}, \mathbf{Y}^{[g]})$ where $\mathbf{Y}^{[g]} = (\mathbf{Y}^{[g^m]'}, \mathbf{Y}^{[g^o]'})'$ 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 needs 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.

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

$$\mathbf{Y}^{[g^m]} | \mathbf{Y}^{[g^o]}, \mathcal{H} \sim t_{m \times gp}(\mu_{(u|g)}, \boldsymbol{\Phi}_{(u|g)} \otimes \boldsymbol{\Psi}_{(u|g)}, \delta_{(u|g)}), \qquad (5.1)$$

where

$$\mu_{(u|g)} = \mu_{(1)} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{Y}^{[g^o]} - \mu_{(2)}) \qquad (5.2)$$

$$\boldsymbol{\Phi}_{(u|g)} = \frac{\delta_1 - gp + 1}{\delta_1 - gp + n - m + 1}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}) \qquad (5.3)$$

$$\boldsymbol{\Psi}_{(u|g)} = \frac{1}{\delta_1 - gp + 1}\{\boldsymbol{\Lambda}_1 \otimes \boldsymbol{\Omega} + (\mathbf{Y}^{[g^o]} - \mu_{(2)})'\mathbf{A}_{22}^{-1}(\mathbf{Y}^{[g^o]} - \mu_{(2)})\} \qquad (5.4)$$

$$\delta_{(u|g)} = \delta_1 - gp + n - m + 1 \qquad (5.5)$$

(Le & Zidek, 2006, p.160–161).

To obtain the one–step–ahead temporal prediction at the 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

$$\mathbf{Y}^{[g^m]} = ((Y_{121,1}^{[g_1^m]}, \ldots, Y_{121,p}^{[g_g^m]})', \ldots, (Y_{120,1}^{[g_1^o]}, \ldots, Y_{120,p}^{[g_g^o]})')',$$

with $\mathbf{Y}_{121,1:p}^{[g_{1:g}^m]} : 1 \times gp$, the unobserved response vector of day 121 and $\mathbf{Y}_{120,1:p}^{[g_{1:g}^o]} : 1 \times gp$, the observed response vector of day 120. Hence $m = 2$ and $n = 121$. The predictive posterior distribution of $\mathbf{Y}^{[g^m]}$ can be obtained by using (5.2)–(5.5). To obtain the predictive distribution of $\mathbf{Y}_{121,1:p}^{[g_{1:g}^m]}$ given $\mathbf{Y}_{1:119,1:p}^{[g_{1:g}^o]}$ and $\mathbf{Y}_{120,1:p}^{[g_{1:g}^o]}$, one can decompose $\mu_{(u|g)}$, $\mathbf{\Phi}_{(u|g)}$ and $\mathbf{\Psi}_{(u|g)}$ as follows:

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

and

$$\delta_{(u|g)}\mathbf{\Phi}_{(u|g)} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix} : \begin{pmatrix} 1 \times 1 & 1 \times 1 \\ 1 \times 1 & 1 \times 1 \end{pmatrix}.$$

Hence, the predictive posterior distribution of $\mathbf{Y}_{121,1:p}^{[g_{1:g}^m]}$ is given by

$$\begin{aligned}
\mathbf{Y}_{121,1:p}^{[g_{1:g}^m]} | \mathbf{Y}_{120,1:p}^{[g_{1:g}^o]}, \mathbf{Y}_{1:119,1:p}^{[g_{1:g}^o]}, \mathcal{H} \quad \sim \quad & t_{1 \times gp}(\mu_{1r} + \mathbf{B}_{12}\mathbf{B}_{22}^{-1}(\mathbf{Y}_{120,1:p}^{[g_{1:g}^o]} - \mu_{2r}), \\
& \frac{(\mathbf{B}_{11} - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21})}{\delta_{(u|g)} + 1} \otimes \mathbf{\Psi}_{(u|g)}(\mathbf{I}_{gp} \\
& + \mathbf{\Psi}_{(u|g)}^{-1}(\mathbf{Y}_{120}^{[g_{1:g}^o]} - \mu_{2r})'\mathbf{B}_{22}^{-1}(\mathbf{Y}_{120}^{[g_{1:g}^o]} \\
& - \mu_{2r}))). \hspace{3cm} (5.6)
\end{aligned}$$

Let $\mathbf{E}_1 : gp \times g$ be a block diagonal matrix with the block $\mathbf{e}_p : p \times 1$, having p^{th} diagonal element 1 and all others 0. At Gauged Site $j \in \{1, \ldots, g\}$, the predictive distribution of the pth unobserved response $\mathbf{Y}_{121,p}^{[g_j^m]}$, that is, $\mathbf{Y}_{121,1:p}^{[g_{1:g}^m]}\mathbf{E}_1\mathbf{e}_j$, also has a t–distribution:

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

$$(5.7)$$

where $\mu^* = \mu_{1r} + \mathbf{B}_{12}\mathbf{B}_{22}^{-1}(\mathbf{Y}_{120,1:p}^{[g_{1:g}^o]} - \mu_{2r})$, $\phi^* = (\mathbf{B}_{11} - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21})/(\delta_{(u|g)}$
$+1)$ and $\mathbf{\Psi}^* = \mathbf{\Psi}_{(u|g)}(\mathbf{I}_{gp} + \mathbf{\Psi}_{(u|g)}^{-1}(\mathbf{Y}_{120,1:p}^{[g_{1:g}^o]} - \mu_{2r})'\mathbf{B}_{22}^{-1}(\mathbf{Y}_{120,1:p}^{[g_{1:g}^o]} - \mu_{2r}))$.

- 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 ungauged sites in the field, one has $m = 1$ and $n = 120$ in (5.2)–(5.5). Let $\mathbf{E}_{2j} : gp \times p$ be the matrix obtained by stacking g $(p \times p)$ matrices, in which the jth matrix is $\mathbf{I}_p$ and others are 0. At Gauged Site $j \in \{1, \ldots, g\}$, we have

$$\mathbf{Y}^{[g^m]}\mathbf{E}_{2j}|\mathbf{Y}^{[g^o]}, \mathcal{H} \quad \sim \quad t_{1 \times p}(\mu_{(u|g)}\mathbf{E}_{2j}, \mathbf{\Phi}_{(u|g)} \otimes \mathbf{E}_{2j}'\mathbf{\Psi}_{(u|g)}\mathbf{E}_{2j}, \delta_{(u|g)}).$$
$$(5.8)$$

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 $\mathbf{E}_3 :$ $p \times p$ be such a matrix that its entries in the ith row and $(p - i + 1)^{\text{th}}$ column are 1 while all others, 0, for $i = 1, \ldots, p$. 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 locating at the first position of the new response vector, the response of the second last hour locating at 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:

$$\mathbf{Y}^{[g^m]}\mathbf{E}_{2j}\mathbf{E}_3|\mathbf{Y}^{[g^o]}, \mathcal{H} \quad \sim \quad t_{1 \times p}(\mu_j, \mathbf{\Phi}_{(u|g)} \otimes \mathbf{\Psi}_j, \delta_{(u|g)}), \quad (5.9)$$

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

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

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

$$
\begin{aligned}
\mathbf{T}_{1c}|\mathbf{T}_{2c}, \mathbf{Y}^{[g^o]}, \mathcal{H} \quad \sim \quad & t_{1\times(k-1)}\big(\mu_{1c} + (\mathbf{T}_{2c} - \mu_{2c})\mathbf{C}_{22}^{-1}\mathbf{C}_{21}, \frac{\delta_{(u|g)}}{\delta_{(u|g)} + p - k + 1} \\
& \times \boldsymbol{\Phi}_{(u|g)}\{1 + (\delta_{(u|g)}\boldsymbol{\Phi}_{(u|g)})^{-1}(\mathbf{T}_{2c} - \mu_{2c})\mathbf{C}_{22}^{-1} \\
& \times (\mathbf{T}_{2c} - \mu_{2c})'\} \otimes (\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}), \\
& \delta_{(u|g)} + p - k + 1\big). \qquad\qquad\qquad (5.10)
\end{aligned}
$$

The last "pollutant", that is, the first entry of $\mathbf{T}_{1c}$, can be predicted by multiplying $\mathbf{T}_{1c}$ with a $(k-1)$–dimensional vector E_4, in which its first entry is 1 and 0 otherwise. Consequently, the predictive posterior distribution of $Y_{n,k-1}^{[g_j^m]}$ is given as follows:

$$
\begin{aligned}
Y_{n,k-1}^{[g_j^m]}|\mathbf{Y}^{[g^o]}, \mathbf{Y}_{n-1,k:p}^{[g_j^o]}, \mathcal{H} \quad \sim \quad & t_{\delta_{(u|g)}+p-k+1}\big((\mu_{1c} + (\mathbf{T}_{2c} - \mu_{2c})\mathbf{C}_{22}^{-1}\mathbf{C}_{21})\mathbf{E}_4, \\
& \frac{\delta_{(u|g)}}{\delta_{(u|g)} + p - k + 1}\boldsymbol{\Phi}_{(u|g)}\{1 \\
& + (\delta_{(u|g)}\boldsymbol{\Phi}_{(u|g)})^{-1}(\mathbf{T}_{2c} - \mu_{2c})\mathbf{C}_{22}^{-1} \\
& \times (\mathbf{T}_{2c} - \mu_{2c})'\}\mathbf{E}_4'(\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21})\mathbf{E}_4\big).
\end{aligned}
$$

$$(5.11)$$

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

$$
\begin{aligned}
\mathrm{Var}(Y_{n,k-1}^{[g_j^m]})|\mathbf{Y}^{[g^o]}, \mathbf{Y}_{n-1,k:p}^{[g_j^o]}, \mathcal{H}) \quad = \quad & \frac{\delta_{(u|g)}}{\delta_{(u|g)} + p - k - 1}\boldsymbol{\Phi}_{(u|g)}\{1 \\
& + (\delta_{(u|g)}\boldsymbol{\Phi}_{(u|g)})^{-1}(\mathbf{T}_{2c} - \mu_{2c})\mathbf{C}_{22}^{-1} \\
& \times (\mathbf{T}_{2c} - \mu_{2c})'\}\mathbf{E}_4'(\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1} \\
& \mathbf{C}_{21})\mathbf{E}_4.
\end{aligned}
$$

It is straightforward to construct the 95% pointwise predictive intervals at the $(k-1)^{\text{th}}$ hour of the 121st day at each gauged site from (5.11).

5.2 The DLM Approach

An alternative approach, *DLM*, can also be used for temporal prediction and indeed would seem the obvious choice, it being an amalgamation of state–space time series models. For the measurement and state equations of the DLM

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

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, T, that is, $\mathbf{x}_T|\mathbf{y}_{1:T}, \theta \sim N(\mathbf{m}_T, \sigma^2 \mathbf{C}_T)$, using the Kalman filter, a smoothing method and the Metropolis–within–Gibbs sampling algorithm (details in Chapter 2).

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

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

$$(5.12)$$

for $h \in \mathcal{N}$. Hence, $T = 2880$ and $h = 1, \ldots, 24$ for the one–day ahead prediction. For any fixed h, the predictive response, $\mathbf{y}_{T+h}$, can also be obtained by the MCMC 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 algorithm. That is, one has

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

Then, the predictive response at iteration j, $\mathbf{y}_{T+h}^{(j)}$, can be drawn from

(5.12), that is,

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

Consequently, the predictive responses are obtained by the sample means of $\{\mathbf{y}_{T+h}^{(j)} : j = 1, \ldots, J\}$ ($J = 500; h = 1, \ldots, 24$). The empirical predictive intervals at the 95% nominal level are obtained by the quantiles of these samples.

5.3 NAIVE* Approach

The other alternative approach, NAIVE*, helps us assess the one–day ahead prediction performance of the multivariate BSP and DLM approaches. 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 of the gauged sites. Using the same notation as above, at each gauged site $j \in \{1, \ldots, g\}$, the response variable is $\mathbf{Y}_{1:n}^j = (Y_{1,1}^{[g_j^o]}, \ldots, Y_{1,p}^{[g_j^o]}, \ldots, Y_{n,1}^{[g_j^o]}, \ldots, Y_{n,p}^{[g_j^o]})' : 1 \times np$, 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 day effect of the week, that is, Monday, Tuesday, etc.; and the last for the hours, a substitute for day effect but giving more reasonable results. In other words, its "design matrix" can be written as

$$
\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 1 & p \\ \vdots & \vdots & \vdots \\ 1 & n & 1 \\ \vdots & \vdots & \vdots \\ 1 & n & p \end{pmatrix} : (np) \times 3.
$$

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, $\widehat{\beta_j} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}^j_{1:n}$.

The one–day ahead prediction at Gauged Site j is thus given by $\widehat{\mathbf{Y}}^j_{n+1} = \mathbf{X}_{n+1}\widehat{\beta}_j$, where $\mathbf{X}_{n+1} : p \times 3$ as follows:

$$
\mathbf{X}_{n+1} = \begin{pmatrix} 1 & n+1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & n+1 & p \end{pmatrix}.
$$

Next the multivariate BSP approach is compared to the DLM and NAIVE* for one–day–ahead prediction at gauged sites in the Chicago area.

5.4   Results and Comparisons

Figures 5.1– 5.14 plot the temporal predictions of the square–root of ozone levels on the 121st day by the multivariate BSP, univariate DLM and NAIVE* approaches, the 95% pointwise predictive 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 multivariate BSP is much more accurate than either the DLM or NAIVE* approaches. In fact, its predictive performance is rather good at most gauged sites.

Table 5.1 presents the mean square predictive error (MSPE) of the predictive responses on the 121st day at each one of the 14 gauged sites using the three approaches. At Gauged Site j, the MSPE of the prediction at hour h can be computed by:

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

where PRED^j_h is the predictive response at hour h of the 121st day and OBS^j_h, the observed response at the same hour of the 121st day, at Gauged Site j. The DLM has the poorest MSPE over all the gauged sites compared with NAIVE* and the BSP. The NAIVE* approach is slightly better than

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 5.1: The mean square predictive error (MSPE) of the one–day ahead prediction at the 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.

the BSP at Gauged Sites 5, 6, and 10. The BSP has the smallest MSPE across most gauged sites among these three.

Figure 5.15 plots the length of the 95% pointwise predictive intervals by the BSP at the 24 hours of the 121st day. Starting from the middle hours of that day, i.e., 9 A.M., the predictive 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 5.16 plots the length of the empirical 95% predictive intervals by the DLM at the 24 hours of the 121st day. These lengths are close to each other but have the wiggly periodic behaviour across all gauges sites, a characteristic previously observed in Chapter 2. Though these lengths are very close to each other, the DLM actually underestimates the predictive variabilities at the gauged sites as seen in Figure 5.17 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 5.1 can be straightforward generalized to an arbitrary time point, not limiting to the case of 121 days of response vectors in this chapter.

Therefore, we conclude that the multivariate BSP approach is more accurate on the one–day ahead prediction at the gauged sites in the Chicago area AQS 2000 database than both the NAIVE and DLM approaches.
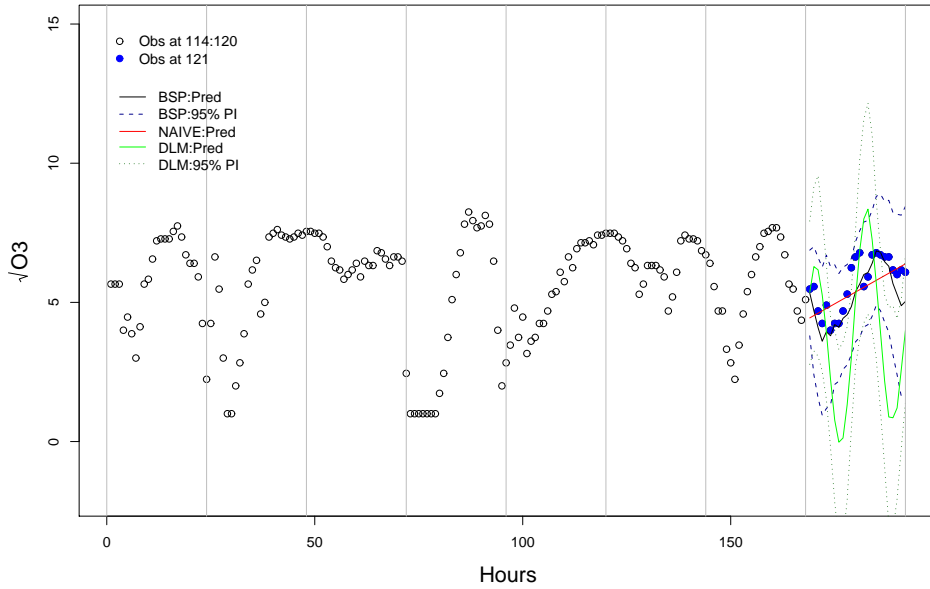
Figure 5.1: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from day 114 to day 121, the predicted values using the multivariate BSP, DLM and NAIVE* approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Gauged Site 1.

5.5   Conclusion and Discussion

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 adjusted approach can be generalized to any time point other than 121 in
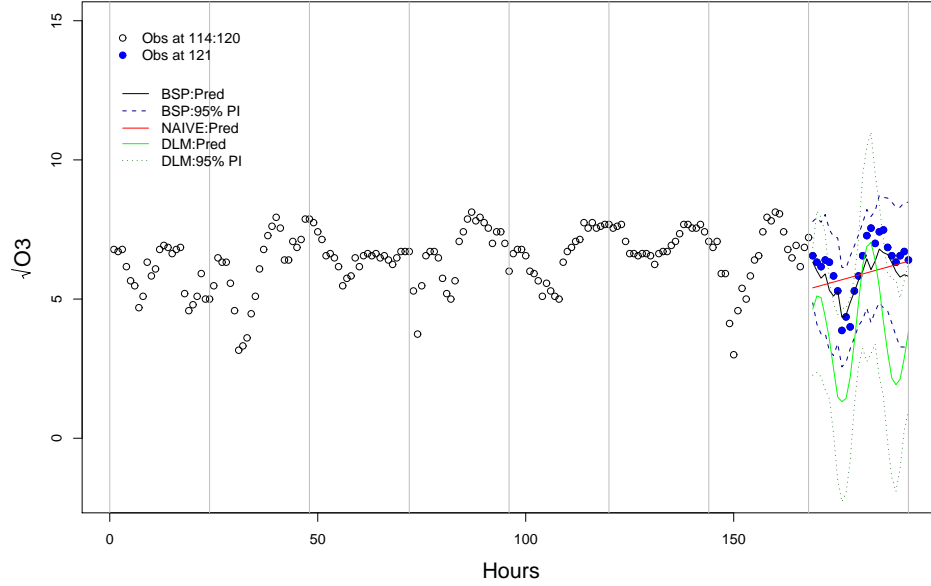
Figure 5.2: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from day 114 to day 121, the predicted values using the multivariate BSP, DLM and NAIVE* approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Gauged Site 2.

the above analysis. That extension is straightforward and will be given in a future work. Moreover, forecasting more hours will be feasible when more data are available. This approach could provide a whole new approach to univariate time series.

The potential problem with this adjusted approach is due to the loss of the information when only a subset of the whole database is used. Further extensions of the correlated response vector in the multivariate BSP modelling need to be explored. One possible solution is to develop the dynamic version of the multivariate BSP, the topic for the next chapters.
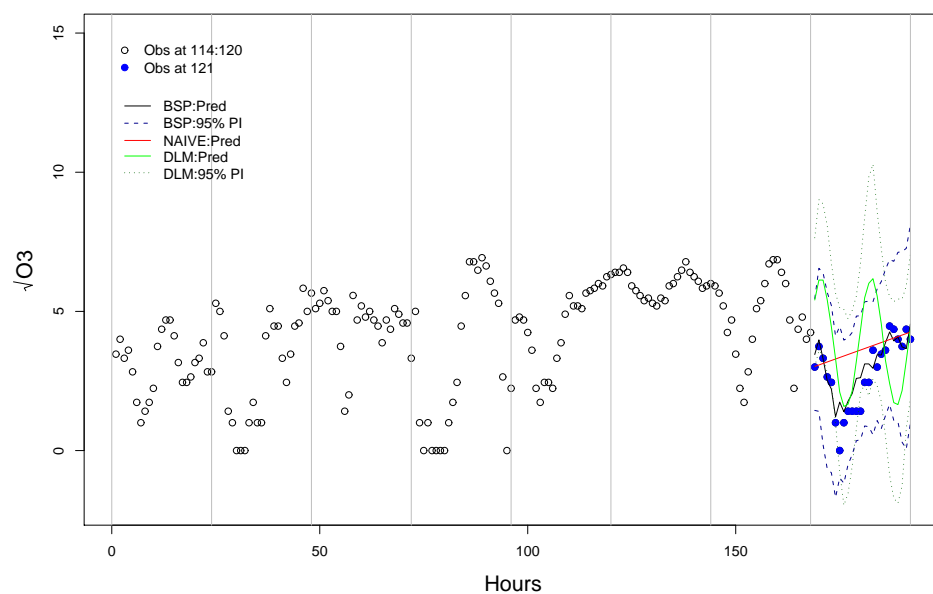
Figure 5.3: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from day 114 to day 121, the predicted values using the multivariate BSP, DLM and NAIVE[*] approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Gauged Site 3.
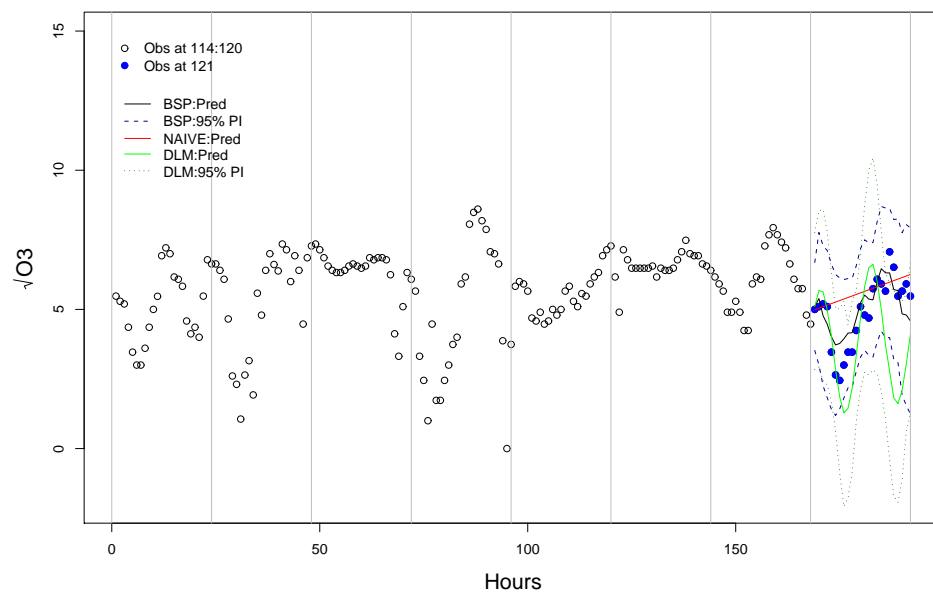
Figure 5.4: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from day 114 to day 121, the predicted values using the multivariate BSP, DLM and NAIVE* approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Gauged Site 4.
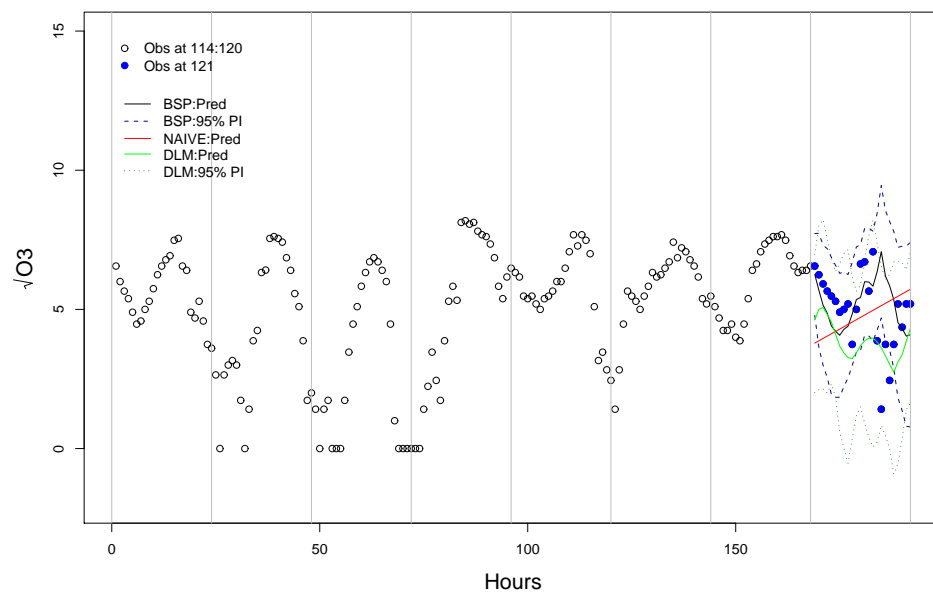
Figure 5.5: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from day 114 to day 121, the predicted values using the multivariate BSP, DLM and NAIVE* approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Gauged Site 5.
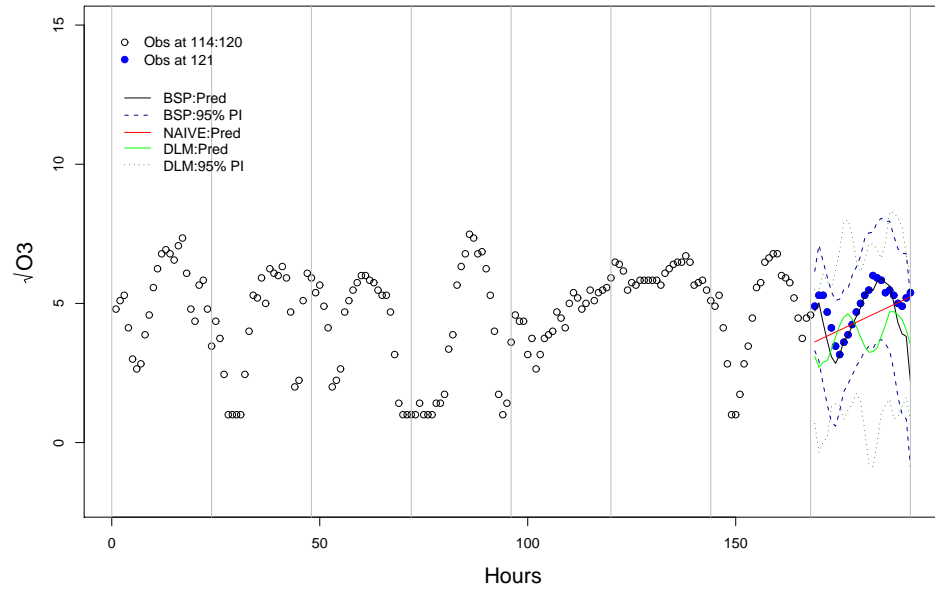
Figure 5.6: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from day 114 to day 121, the predicted values using the multivariate BSP, DLM and NAIVE* approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Gauged Site 6.
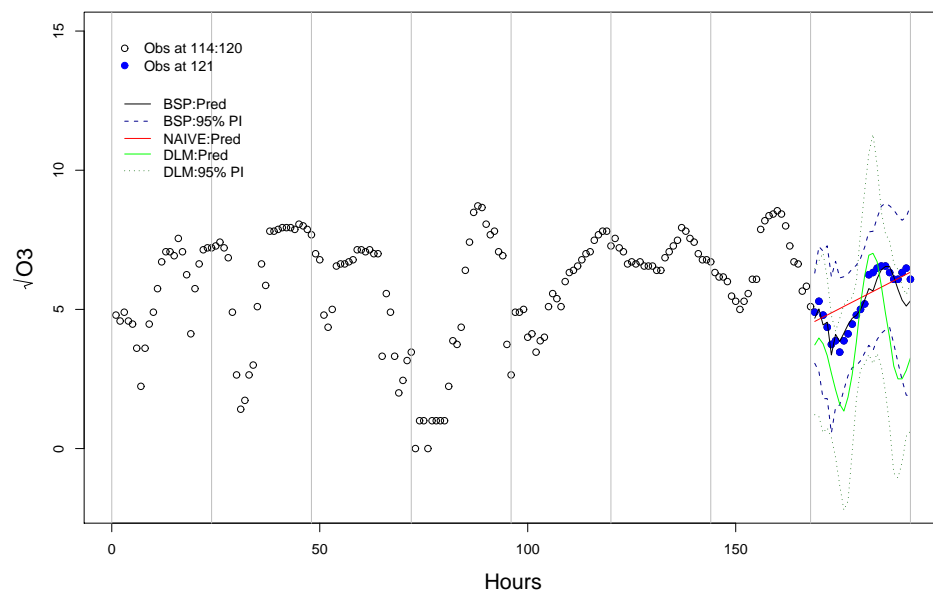
Figure 5.7: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from day 114 to day 121, the predicted values using the multivariate BSP, DLM and NAIVE* approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Gauged Site 7.
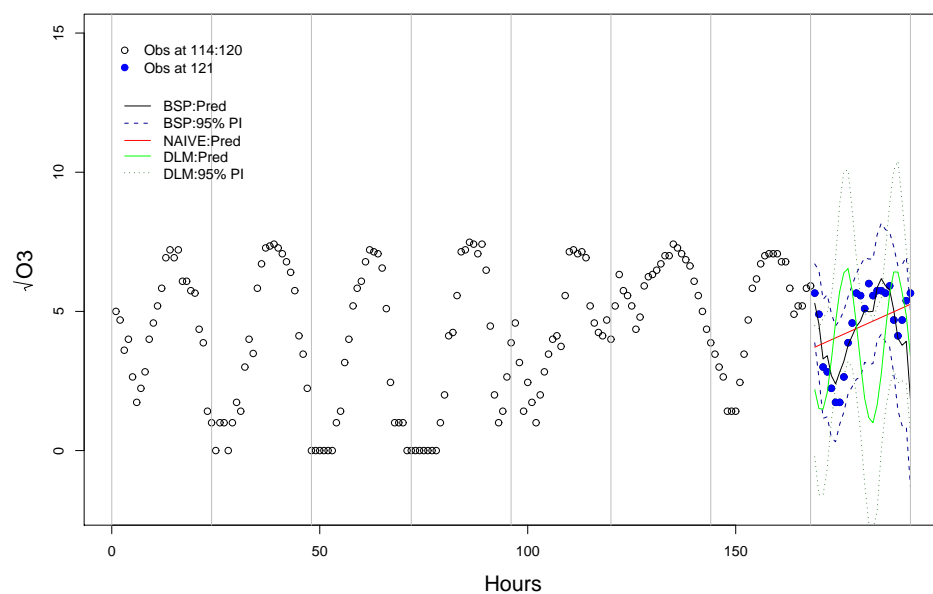
Figure 5.8: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from day 114 to day 121, the predicted values using the multivariate BSP, DLM and NAIVE* approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Gauged Site 8.
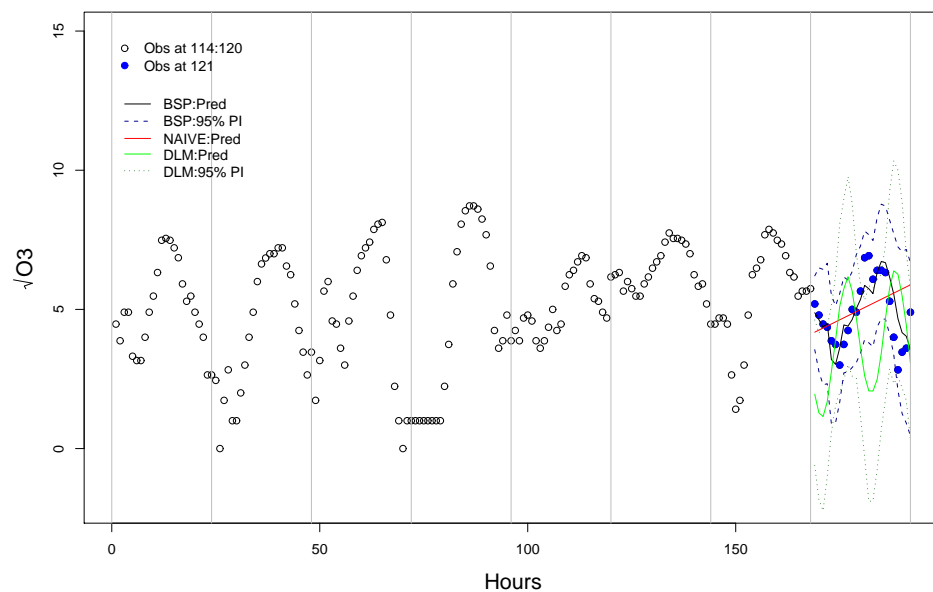
Figure 5.9: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from day 114 to day 121, the predicted values using the multivariate BSP, DLM and NAIVE* approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Gauged Site 9.

117
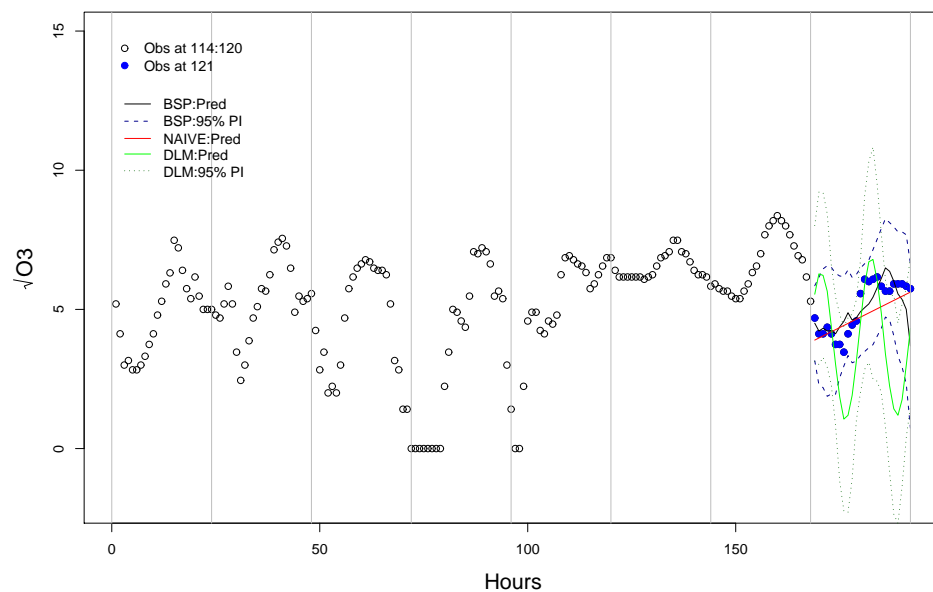
Figure 5.10: The observed square–root of ozone concentrations $(\sqrt{\text{ppb}})$ from day 114 to day 121, the predicted values using the multivariate BSP, DLM and NAIVE* approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Gauged Site 10.
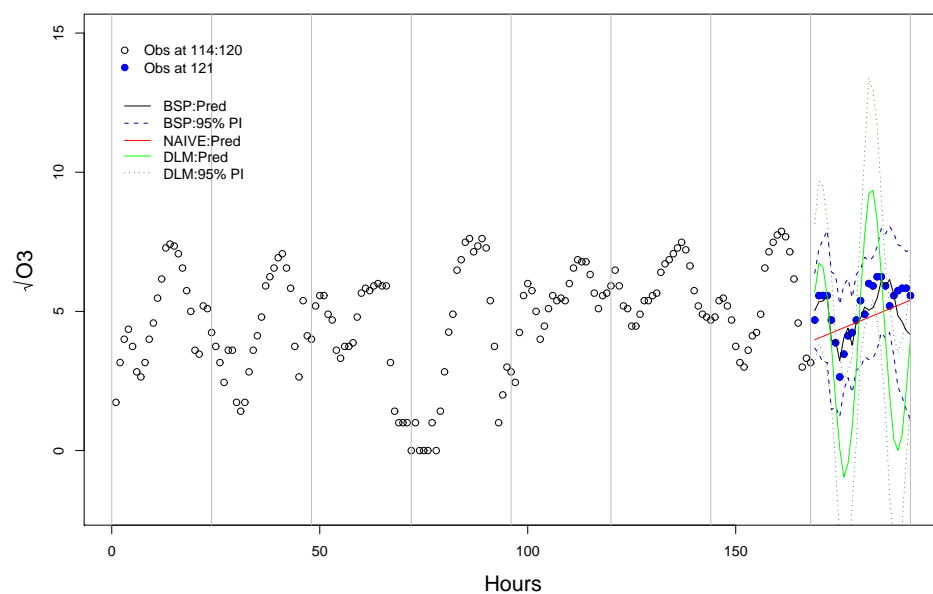
Figure 5.11: The observed square–root of ozone concentrations $(\sqrt{\text{ppb}})$ from day 114 to day 121, the predicted values using the multivariate BSP, DLM and NAIVE* approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Gauged Site 11.
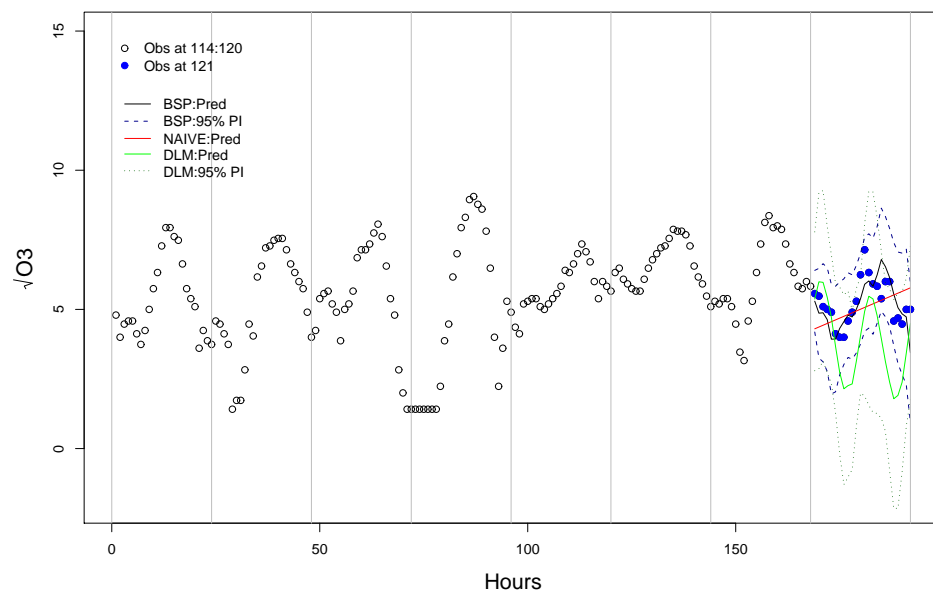
Figure 5.12: The observed square–root of ozone concentrations $(\sqrt{\text{ppb}})$ from day 114 to day 121, the predicted values using the multivariate BSP, DLM and NAIVE* approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Gauged Site 12.
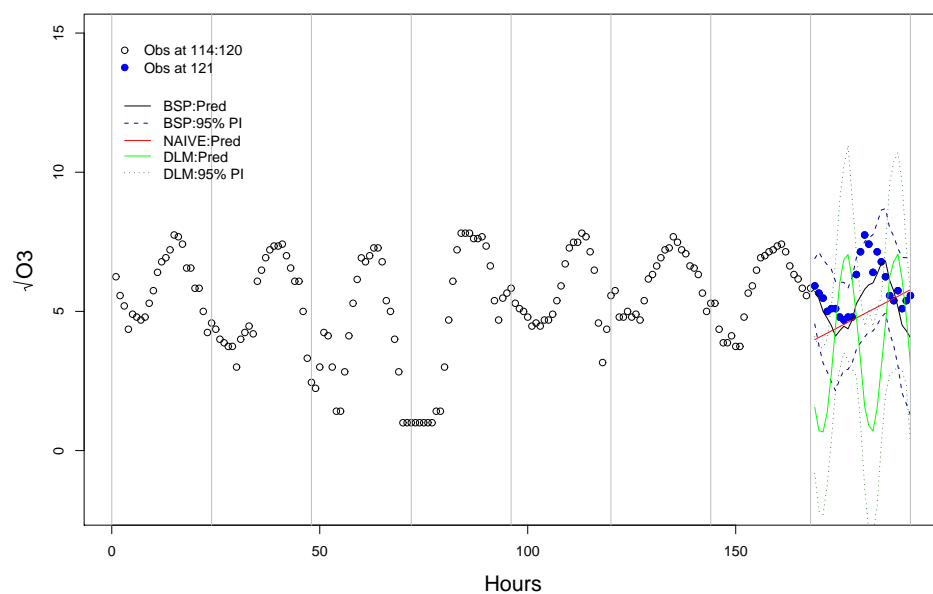
Figure 5.13: The observed square–root of ozone concentrations ($\sqrt{\text{ppb}}$) from day 114 to day 121, the predicted values using the multivariate BSP, DLM and NAIVE[*] approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Gauged Site 13.
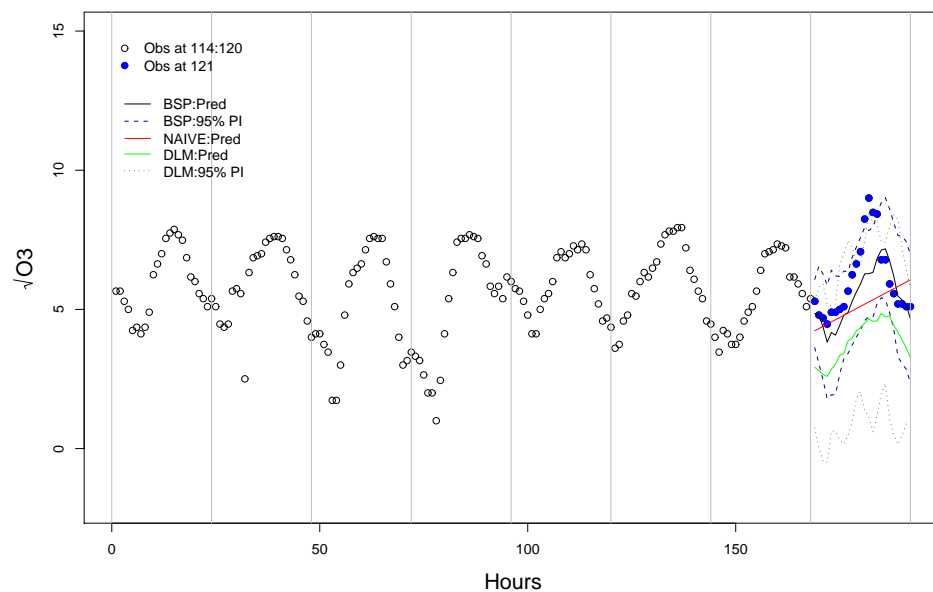
Figure 5.14: The observed square–root of ozone concentrations $(\sqrt{\text{ppb}})$ from day 114 to day 121, the predicted values using the multivariate BSP, DLM and NAIVE* approaches, and the 95% pointwise predictive intervals using the multivariate BSP and DLM approaches at Gauged Site 14.
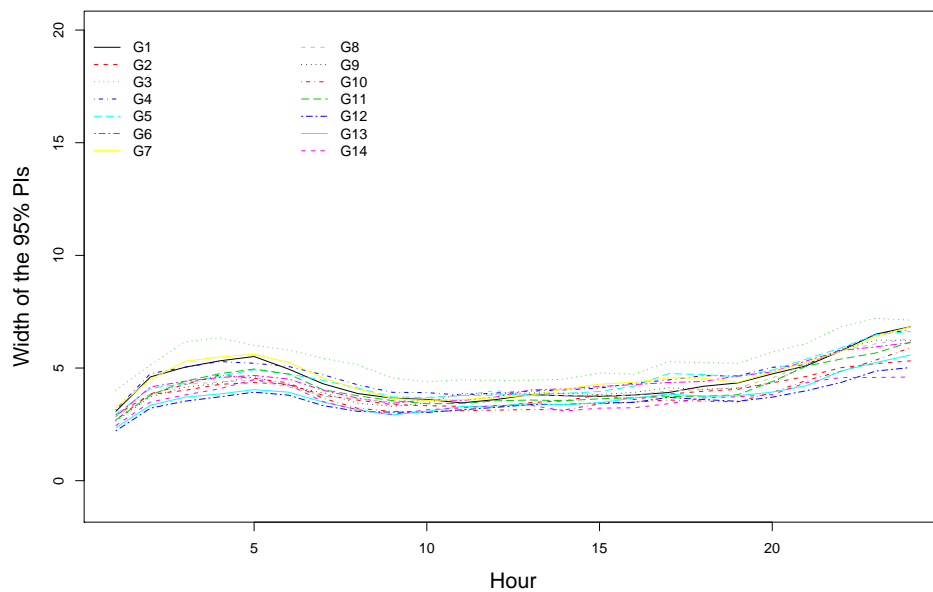
Figure 5.15: The width of the 95% pointwise predictive intervals of the one–day ahead prediction at the 14 gauged sites using the multivariate BSP approach.
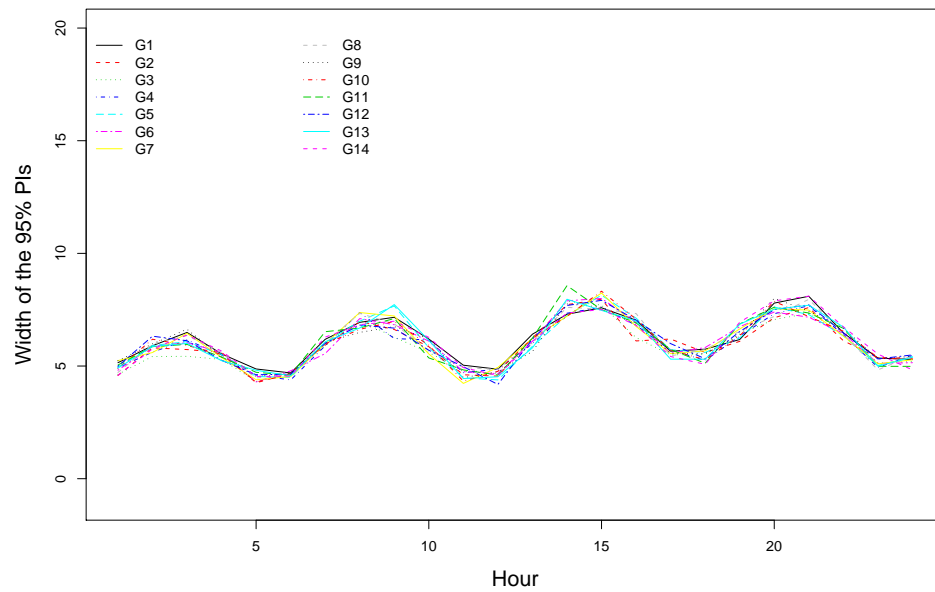
Figure 5.16: The width of the 95% pointwise predictive intervals of the one–day ahead prediction at the 14 gauged sites using the DLM approach.
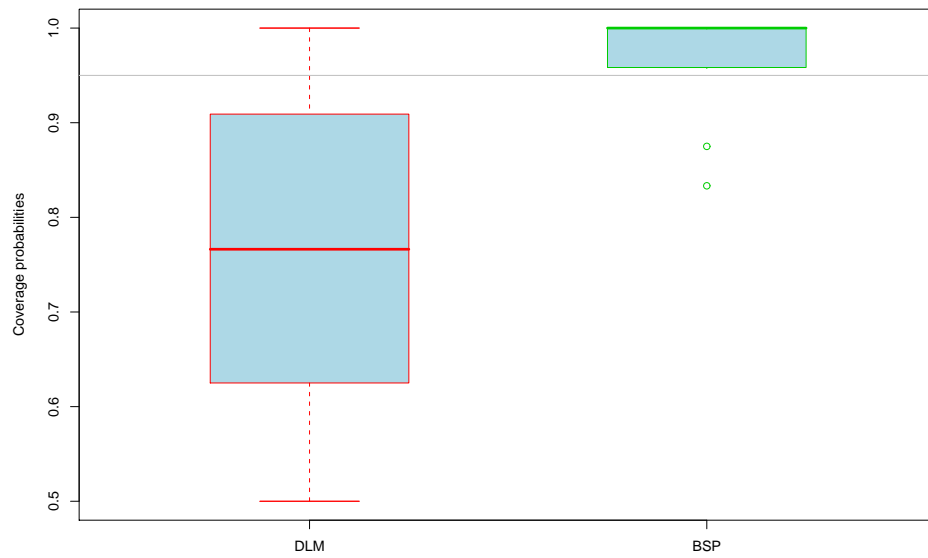
Figure 5.17: Boxplots of the coverage probabilities using the DLM and multivariate BSP approaches at the 95% nominal level.

Chapter 6

Bayesian Empirical
Orthogonal Function Method

This chapter contains some theoretical results relevant to the extension of the BSP approach in next chapter. The idea is straightforward in that we partition the spatial variation into long–term variation, and short–term variation including measurement errors. We obtain what we call "Bayesian (empirical orthogonal functions) EOFs" in this chapter and extend results to incorporate the very flexible GIW prior to model the staircase patterned data in Chapters 4–5.

Section 6.1 introduces empirical orthogonal functions (EOFs). Section 6.2 describes the classical EOFs obtained from spatio–temporal data and discusses potential difficulties with this method. Section 6.3 illustrates an alternative, corrected EOFs, for a known temporal covariance in the separable space–time covariance structure. Section 6.4 proposes a Bayesian EOF method for a general Bayesian hierarchical model. Section 6.5 extends those results to incorporate the GIW prior. Simulation Study 1 is presented in Sections 6.2–6.3. Section 6.6 compares those three types of EOFs by Simulation Study 2. Section 6.7 summarizes the results and states conclusions for this chapter.

6.1 Introduction

Over the years, empirical orthogonal functions (EOFs) have been used extensively for identifying spatial patterns for environmental processes. EOFs implicitly assume that the samples consist of independent replicates of the spatial fields. These have been widely used in astronomy, physics, clima-

tology, and oceanography (see Hannachi et al. (2007) for a review on this subject in atmospheric science).

EOFs deal with data anomalies, that is, deviations of the observed values from their temporal mean, instead of the responses. These anomalies are obtained by subtracting the time averages of the responses from the local responses at each time point and site. More precisely, let $Z_i(t)$ represent the response at site $\mathbf{s}_i$ and at time t, for $i = 1, \ldots, p$ and $t = 1, \ldots, n$. Then the anomaly at site $\mathbf{s}_i$ and time t is defined as

$$Y_i(t) \;\; = \;\; Z_i(t) - \frac{1}{n} \sum_{t=1}^{n} Z_i(t), \tag{6.1}$$

for $i = 1, \ldots, p$ and $t = 1, \ldots, n$.

Wikle (2002) describes EOFs as a commonly used spatio–temporal method in Climatology. The method is widely used in spatial process analysis to detect spatial patterns in a given field. For a continuous spatial process observed at discrete time points, the EOFs represent components of a Karhunen–Loève (KL) expansion; while in the discrete process, they are the components of a Principal Component Analysis (PCA).

EOF and PCA differ from each other. EOF can be applied on any irregularly located monitoring stations to find both the time series and spatial patterns. PCA can also be applied on any sites but only obtain the time series or spatial patterns. The spatial pattern represent the mode of variability. The time series patterns, also called as the time series amplitudes, reflect or characterize how the spatial patterns oscillate over time. The time series amplitudes are often referred as the *expansion coefficients* or *principle components*. In this thesis, we call the time series amplitudes as expansion coefficients, in order to distinguish them from the principal components, the term often used in PCA. Although both EOF and PCA can be applied on any site, EOF might be inaccurate if it is constructed given unevenly located sites.

In Climatology literature, some authors define EOF and PCA differently but some not. We can use infinitely many orthonormal basis vectors to

construct the EOF. However, the eigenvector basis is the only one that allows the expansion coefficients having the PCA property, that is, the expansion coefficients of the EOFs are uncorrelated (Björnsson and Venegas, 1997).

The KL expansion represents a stochastic process by a linear combination of an infinite number of orthogonal functions. In the KL expansion, the coefficients for the orthogonal functions are random variables. The orthogonal functions are the eigenfunctions of the covariance function for this process. For the anomalies data, suppose $\mathbf{Y}_t = (Y_t(\mathbf{s}_1), \ldots, Y_t(\mathbf{s}_p))' : p \times 1$ represents the anomalies vector at t across all the spatial sites in the region. Let $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n) : p \times n$ be the anomaly matrix.

Definition 6.1.1 *(KL expansion) Consider the spatio–temporal process $\{Z(\mathbf{s}, t) : \mathbf{s} \in \mathcal{D}, t = 1, \ldots, n\}$, where $\mathbf{s}$ represents for the spatial location in the domain $\mathcal{D} = \{\mathbf{s}_1, \ldots, \mathbf{s}_p\}$ that of interest while t represents a time point. Assume $E(Z(\mathbf{s}, t)) = 0$ and $Cov(Z(\mathbf{s}_i, t), Z(\mathbf{s}_j, t)) = C_0(\mathbf{s}_i, \mathbf{s}_j)$ for all t. The KL expansion represents the covariance function as an infinite linear combination of orthogonal functions, that is,*

$$C_0(\boldsymbol{u}, \boldsymbol{v}) \;\; = \;\; \sum_{j=1}^{\infty} \lambda_j \phi_j(\boldsymbol{u}) \phi_j(\boldsymbol{v}), \tag{6.2}$$

where $\{\lambda_j : j = 1, \ldots, \infty\}$ are the eigenvalues and $\{\phi_j(.) : j = 1, \ldots, \infty\}$, the orthogonal eigenfunctions.

For the complete set of orthonormal basis functions $\{\phi_j(.) : j = 1, \ldots, \infty\}$, the response can be represented as follows:

$$Z(\mathbf{s}, t) \;\; = \;\; \sum_{j=1}^{\infty} a_j(t) \phi_j(\mathbf{s}), \tag{6.3}$$

where $\mathrm{Var}(a_j(t)) > \mathrm{Var}(a_{j+1}(t))$ for $j = 1, 2, \ldots$, and $\mathrm{Cov}(a_i(t), a_j(t)) = 0$ for $i \neq j$. In discrete case, the KL expansion is simply obtained through the PCA. Then $\{\phi_j(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$ is called the jth EOF and $\{a_j(t) : t = 1, \ldots, n\}$, the expansion coefficients corresponding to the jth EOF. In other words, the KL expansion allows one to represent the process by an infinite set of

separable orthonormal basis functions such that these orthonormal basis functions are optimal in minimizing the mean square variance.

We next introduce the term of classical EOFs.

6.2 Classical EOFs

To find the principal spatial patterns in a spatial–temporal process, one can use the *classical EOFs* [6]. The spatial covariance matrix can be estimated using the samples of observed anomalies.

In the geographical region of interest, $z_t(\mathbf{s})$ represents the univariate response variable at site $\mathbf{s}$ and time t, for $s \in \{\mathbf{s}_1, \ldots, \mathbf{s}_p\}$ and $t = 1, \ldots, n$. Let $\mathbf{Z}_t = (z_t(\mathbf{s}_1), \ldots, z_t(\mathbf{s}_p))' : p \times 1$ be the response vector at time t. Assume the matrix–variate response variable $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n) : p \times n$ that follows a matrix–normal distribution with a separable covariance structure in space and time, that is, $\mathbf{Z} \sim N_{p \times n}(\mathbf{0}, \boldsymbol{\Sigma}_S \otimes \boldsymbol{\Sigma}_T)$, where $\boldsymbol{\Sigma}_S : p \times p$ represents the spatial covariance matrix and $\boldsymbol{\Sigma}_T : n \times n$, the temporal covariance matrix. This separable covariance structure implies no space–time interaction in spatial–temporal processes.

Based on these assumptions, the spatial covariance matrix can be estimated by $\hat{\boldsymbol{\Sigma}}_S = \frac{1}{n}\mathbf{YY}'$, where $\mathbf{Y}$ is the anomaly matrix defined in (6.1). The spectral decomposition theorem implies the existence of a unique decomposition for $\hat{\boldsymbol{\Sigma}}_S$ such that $\hat{\boldsymbol{\Sigma}}_S = \hat{\boldsymbol{\Psi}}\hat{\boldsymbol{\Lambda}}^2\hat{\boldsymbol{\Psi}}'$, where $\hat{\boldsymbol{\Lambda}}^2 = \text{diag}\{\hat{\lambda}_1^2, \ldots, \hat{\lambda}_p^2\}$ with $\hat{\lambda}_1 > \ldots > \hat{\lambda}_p > 0$ being the eigenvalues for $\mathbf{ZZ}'$; each column of $\hat{\boldsymbol{\Psi}}$ is the eigenvector corresponding to the associated eigenvalue. Hence, we represent $\boldsymbol{\Sigma}_S$ as

$$\begin{aligned}
\boldsymbol{\Sigma}_S &= \boldsymbol{\Psi}\boldsymbol{\Lambda}^2\boldsymbol{\Psi}' \\
&= (\boldsymbol{\Psi}\boldsymbol{\Lambda})(\boldsymbol{\Psi}\boldsymbol{\Lambda})' \\
&= \boldsymbol{\Phi}\boldsymbol{\Phi}',
\end{aligned} \qquad (6.4)$$

a form of the KL expansion. We then obtain the classical EOFs from $\hat{\boldsymbol{\Phi}} =$

---

[6] "Classical" EOFs are usual EOFs used in literature in which it ignores temporal components.

$\hat{\boldsymbol{\Psi}}\hat{\boldsymbol{\Lambda}}$.

However, the classical EOFs do not efficiently estimate the population level counter parts in $\boldsymbol{\Sigma}_S$ without temporal independence, an unrealistic assumption in most cases. Moreover, this unrealistic assumption might provide misleading information about the spatial pattern obtained by the classical EOFs. In the following section, we will demonstrate this deficiency in classical EOFs through some simulation studies. Moreover, a potential problem arises about how one can avoid the negative effects caused by correlated samples, that is, temporally correlated sequences.

Next in Simulation Study 1, we show severe problems that classical EOFs can cause in a separable space–time field with known EOF matrix. To do this, we need to construct the orthogonal matrix $\mathbf{O}$ in the spectral decomposition theorem. Gram–Schmidt's process is a procedure to obtain an orthogonal basis. The known EOFs can be constructed using the Gram–Schmidt's process to have a specified diagonal matrix $\boldsymbol{\Lambda}$ and an orthonormal basis function $\boldsymbol{\Phi}$. This construction starts with any set of given orthogonal vectors, saying, $\mathbf{O}_1, \ldots, \mathbf{O}_k$, for $1 \leq k \leq n - 1$, and $k \in \mathcal{Z}$. Lemma 6.2.1 gives the details.

Lemma 6.2.1 *Given the orthogonal vectors $\boldsymbol{G}_1, \ldots, \boldsymbol{G}_k$, with $\boldsymbol{G}_j : p \times 1$ for $j = 1, \ldots, k$, we obtain an orthogonal matrix $\boldsymbol{G} = (\boldsymbol{G}_1, \ldots, \boldsymbol{G}_p) : p \times p$ by repeating steps (i)–(iii) for $j = k + 1, \ldots, p$:*

   (i) *Generate a realization $\boldsymbol{y}_j$ from $N_p(\boldsymbol{0}, \boldsymbol{I}_p)$.*

   (ii) *Fit the linear regression model:*

$$\boldsymbol{y}_j = \boldsymbol{A}_0 + \sum_{i=1}^{j-1} \boldsymbol{A}_i \boldsymbol{G}_i + \varepsilon_i,$$

   and obtain the estimated coefficients $\{\hat{\boldsymbol{A}}_i : i = 0, \ldots, j - 1\}$.

   (iii) *Denote $\boldsymbol{G}_j$ to be the fitted residuals $\boldsymbol{y}_j - \hat{\boldsymbol{A}}_0 - \sum_{i=1}^{j-1} \hat{\boldsymbol{A}}_i \boldsymbol{G}_i$ such that $\boldsymbol{G}_j \perp \{\boldsymbol{G}_1, \ldots, \boldsymbol{G}_{j-1}\}$.*

This lemma gives us the orthonormal basis function by normalizing the generated orthogonal vectors $\mathbf{G}_1, \ldots, \mathbf{G}_p$. We hereafter use $\mathbf{O}$ to represent the orthonormal basis matrix using the Gram–Schmidt type expansion. Since $\mathbf{\Sigma}_S = \mathbf{O}\mathbf{\Lambda}^2\mathbf{O}'$, we then obtain the spatial covariance function $\mathbf{\Sigma}_S$ using the above constructed EOFs.

Simulation study 1

The objective of this study is to compare three different types of EOFs for a separable state–space process with known EOFs and known temporal covariance matrix.

Simulated data

In this study, the spatial region is 18×18 grid locations, grid cell edges being taken to be 6 km. The regular lattice–base grids have 324 grid locations. The spatial–temporal process over this lattice is assumed to be a mean 0 process with a separable spatio–temporal covariance structure. After constructing the EOFs matrix using the Gram–Schmidt's method with a known diagonal matrix, the spatial covariance function is formed by their products as presented by the spectral decomposition theorem. The temporal covariance function is simply a causal and invertible AR(1) process with the variance parameter σ_v^2 and AR coefficient, ϕ. The autocorrelation function (ACF) for the AR(1) process is given by

$$\gamma(h) = \begin{cases} \frac{\sigma_v^2}{1-\phi^2} & h = 0 \\ \frac{\sigma_v^2}{1-\phi^2}\phi^{||h||} & h \geq 1. \end{cases}$$

In this example, we choose two orthogonal vectors $\mathbf{G}^{(1)}$ and $\mathbf{G}^{(2)}$ to construct the EOFs using the Gram–Schmidt method. More specifically, for $p = 4q, q = 1, 2, \ldots$, $\mathbf{G}^{(i)}$ is a p–dimensional vector in which its jth entry, $G_j^{(i)}$, is 1 for $j \in [1, \frac{p}{2}]$; -1 for $j \in [\frac{p}{2} + 1, p]$; and 0 otherwise. To motivate development of our simulation model, we think of surface temperature as our response of interest. This response would be strongly determined by lat-

itude, high in summer in the northern hemisphere while low in the southern. We choose $\mathbf{G}_1$ to represent that spatial pattern. We then choose $\mathbf{G}_2$ to represent a High–Low elevation spatial pattern such that $\mathbf{G}_2$ is a p–dimensional vector by repeating $(\mathbf{1}_{\sqrt{p}/2}, -\mathbf{1}_{\sqrt{p}/2})$ [7] for $\sqrt{p}$ times. The constructed orthogonal matrix given by these two orthogonal vectors, denoted by $\mathbf{G}$, can be obtained by applying the Gram–Schmidt method in Lemma 6.2.1.

The first four diagonal entries in the specified diagonal matrix $\mathbf{\Lambda}^2$ are assumed to be 40, 20, 15, and 10. The remaining diagonal entries can be constructed by a decreasing sequence such that their summation is 15 and the minimum or last entry, around 0.023.

Figures 6.1 and 6.2 plot the contours for simulated data at $t = 5$ and $t = 28$ in the region we study, respectively. These graphs show there may be the North–South and High–Low spatial patterns in this field, respectively. Moreover, this spatio–temporal field varies spatially and temporally.

To check the temporal variations in this simulated database, we randomly select four grid locations and plot their histograms, ACFs and PACFs in Figure 6.3. These graphs show a very strong autocorrelation in the time series data of each of the selected four sites, as expected from the strong AR(1) temporal process in the data.

Results and comparisons

We first compute the classical EOFs and compare them with the true EOFs. Figure 6.4 plots the contours for the true EOFs. Clearly the first spatial pattern is North–South spatial pattern and the second, High–Low elevation spatial pattern, the principle determinants of surface temperature. Figure 6.5 plots the contours for the classical EOFs. Although it seems that classical EOFs can also capture the first two types of spatial patterns, obviously the estimates for the contours are far from the truth. For example, the values for the first true EOF in the northern region is close to be 0.35, while for the first classical EOF, close to be 0.6 at the northeastern region and 1.0, the northwestern region. Same sort of things happen for the second true

---

[7]Here $\mathbf{1}_{\sqrt{p}/2}$ represents the $\sqrt{p}/2$–dimensional vector of 1, and so as $-\mathbf{1}_{\sqrt{p}/2}$.
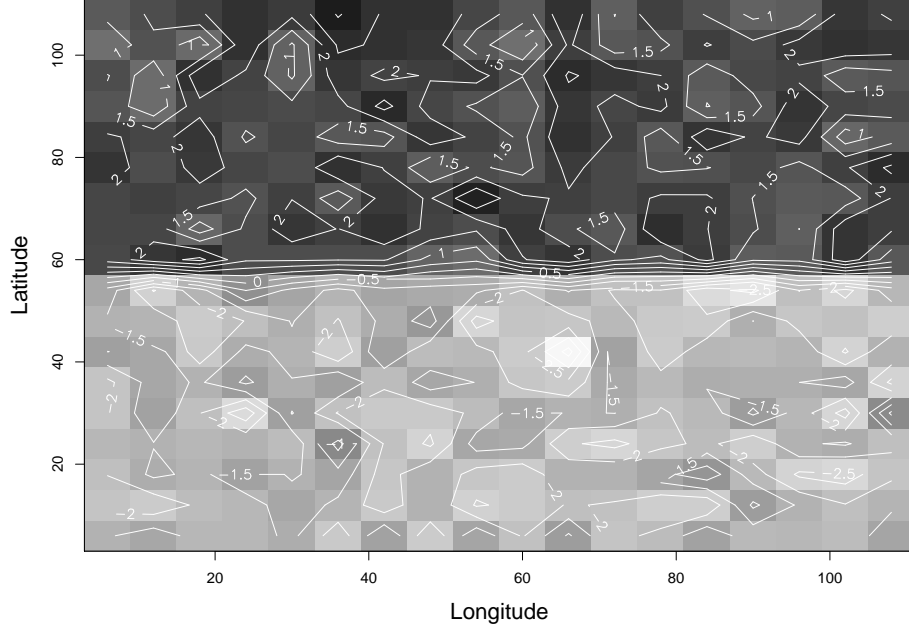
Figure 6.1: Contour plot for the simulated data at day $t = 5$ in the 18×18 grid locations. The AR coefficient in the simulated data is set to be $\phi = 0.9$. (White=-4.0; Black=4.0.)

and classical EOFs. We conclude that the classical EOFs fail to capture the true spatial patterns in this field due to the field's high autocorrelation.

The next section provides an alternative when the temporal covariance matrix, $\boldsymbol{\Sigma}_T$, is known, a method leading to what we call "corrected" EOFs.

6.3 Corrected EOFs

This section presents an alternative to the classical EOFs of Section 6.2, given a known temporal covariance matrix. We call the EOFs obtained *corrected EOFs*, meaning that the temporal dependence structure has been incorporated. Although assuming the temporal covariance $\boldsymbol{\Sigma}_T$ to be known is unrealistic, the analysis exposes the potentially serious problem that the
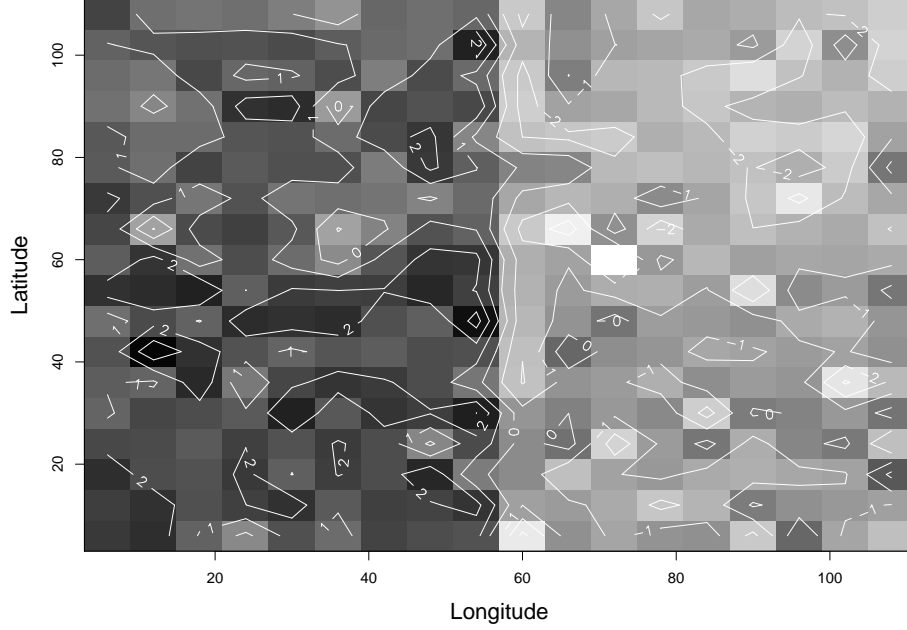
Figure 6.2: Contour plot for the simulated data at day $t = 28$ in the 18×18 grid locations. The AR coefficient in the simulated data is set to be $\phi = 0.9$. (White=-4.0; Black=4.0.)

classical EOFs can cause.

Given the matrix–variate normal distribution for $\mathbf{Y} \sim N_{p \times n}(\mathbf{0}, \boldsymbol{\Sigma}_S \otimes \boldsymbol{\Sigma}_T)$, we have $\mathbf{Y}^* = \mathbf{Y}\boldsymbol{\Sigma}_T^{-1/2} \sim N_{p \times n}(\mathbf{0}, \boldsymbol{\Sigma}_S \otimes \mathbf{I}_n)$, by a standard property of the matrix–variate normal distribution. It is then straightforward to estimate $\boldsymbol{\Sigma}_S$ using $\frac{1}{n} \sum_{t=1}^{n} \mathbf{Y}_t^*(\mathbf{Y}_t^*)' = \frac{1}{n}\mathbf{Y}^*(\mathbf{Y}^*)'$. In other words,

$$\hat{\boldsymbol{\Sigma}}_S \;\; = \;\; \frac{1}{n}\mathbf{Y}\boldsymbol{\Sigma}_T^{-1}\mathbf{Y}', \tag{6.5}$$

given the non–singular covariance matrix $\boldsymbol{\Sigma}_T$. The corrected EOFs are then constructed using the spectral decomposition theorem, same as the classical EOFs. To obtain unique EOFs, we restrict the eigenvectors to form the orthonormal matrix that has positive elements in its first row.

We now revisit the example in Section 6.2 but obtain the corrected EOFs

134
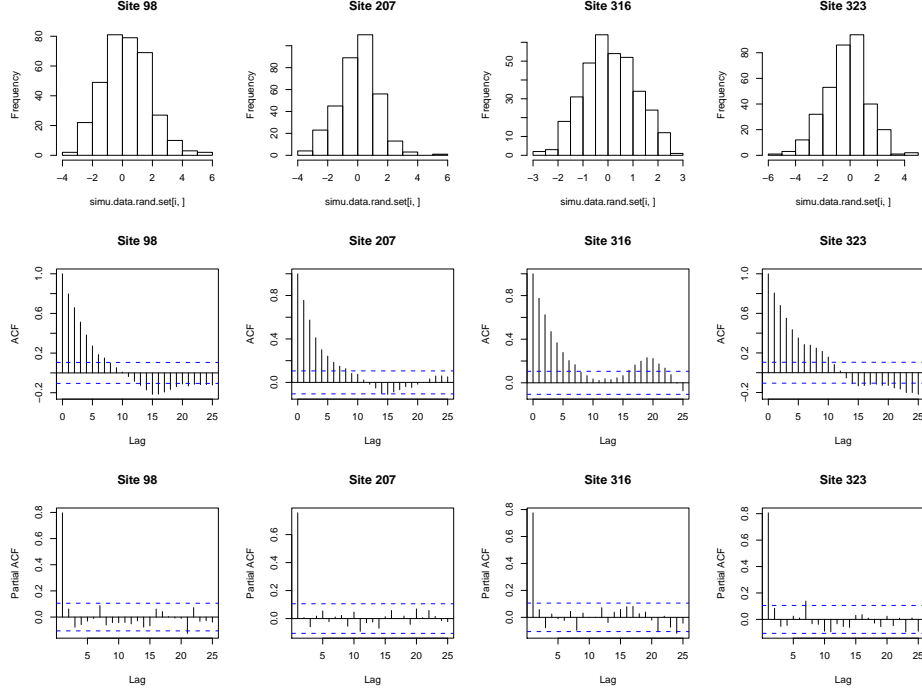
Figure 6.3: Histogram (first row), ACFs (second row) and PACFs (third row) for the simulated data at four randomly selected sites in the region. The AR coefficient in the simulated data is set to be $\phi = 0.9$.

and compare them with the classical and true EOFs to see if the corrected EOFs improve our knowledge about the principal spatial patterns in this simulated database.

Simulation study 1 (revisited)

Figures 6.4–6.5 demonstrate the first two EOFs for true, classical and corrected, respectively. The first two true EOFs seem to have an obvious spatial pattern, seen in the corrected EOFs but not in the classical EOFs. The results demonstrate the potential danger in using the classical EOFs when the spatio–temporal data are highly temporally correlated as well.

Table 6.1 represents the percentage of spatial variation for the first 10 EOFs by the true, classical, and corrected methods. The corrected
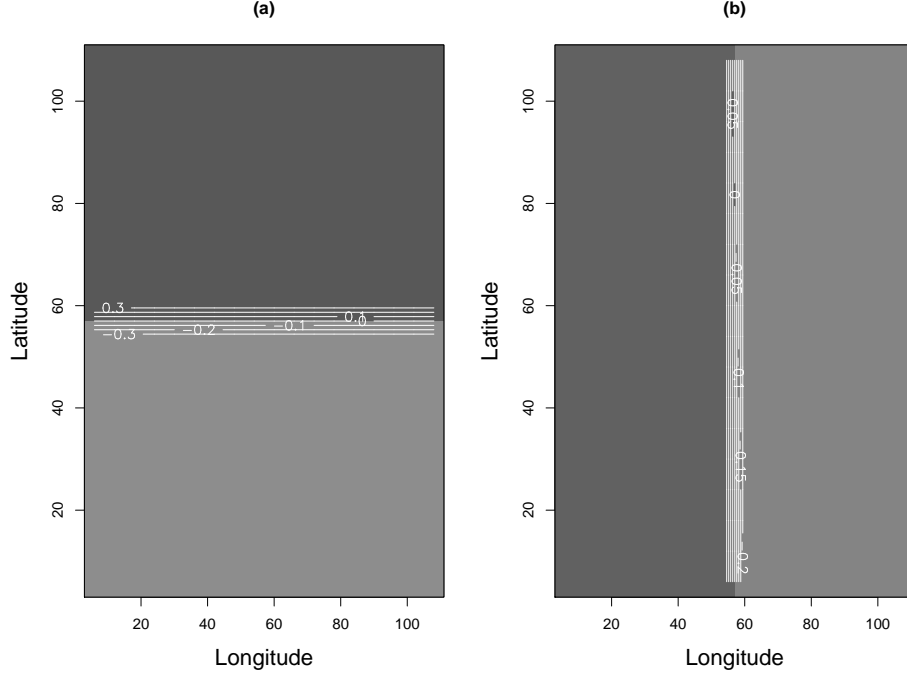
Figure 6.4: Contour plots for the first two true EOF vectors: (a) – 1$^{\text{st}}$ EOF; and (b) – 2$^{\text{nd}}$ EOF. (White=-1.7; Black=1.4.)

EOFs are much closer to the true values than the classical ones. Table 6.2 demonstrates the matrix discrepancies[8] for the classical and corrected EOFs against the true EOFs. Assuming the temporal covariance function is known, the corrected EOFs are closer to the true EOFs than the classical ones. The matrix discrepancy between the true and corrected EOFs is much smaller than that between the true and classical one. Moreover, the corrected EOFs better characterize the first two spatial types in this field than the classical EOFs do.

More generally, for an unknown temporal covariance structure, it is impossible to use the corrected EOFs to identify the principal spatial patterns

---

[8]The *matrix distance* between two matrices, A and B, is defined through the concept of the inner product of these two matrices, that is, $||A - B|| = \text{tr}(AB')$. The *matrix discrepancy* between A and B is defined as $\sqrt{\frac{||A-B||}{||B||}}$.
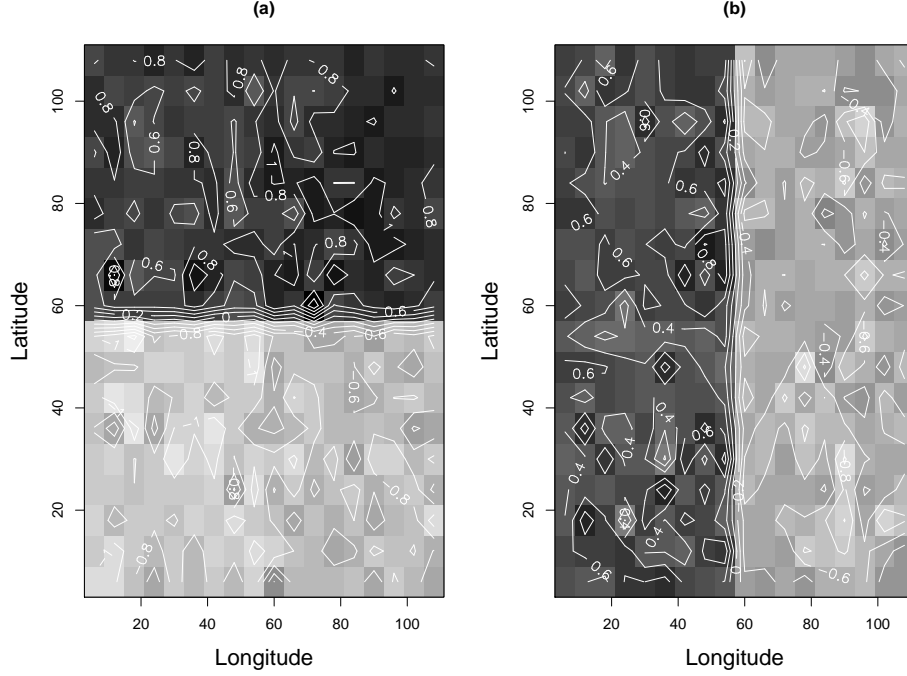
Figure 6.5: Contour plots for the first two classical EOF vectors: (a) – 1$^{\text{st}}$ EOF; and (b) – 2$^{\text{nd}}$ EOF. (White=-1.7; Black=1.4.)

in spatio–temporal fields. One may have interest to know whether we can obtain an accurate estimate of the spatial covariance matrix based on the observations so that the EOFs can better represent the principal spatial patterns. Moreover, how can one take into account of uncertainties in the estimates?

To help answer those questions, we next propose a new Bayesian version of EOFs by adding a prior distribution for the spatial covariance matrix.

6.4 Bayesian EOFs

In practice, the corrected EOFs cannot be computed because of the unknown temporal dependence structure for the spatial–temporal matrix–variate response $\mathbf{Z} : p \times n$, where p is the total number of spatial locations in the do-
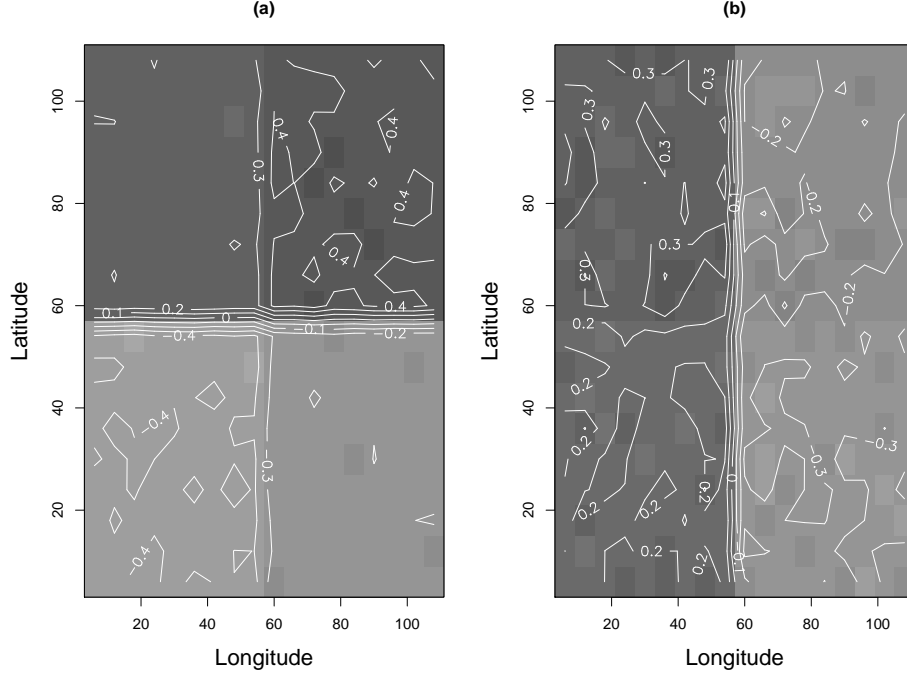
Figure 6.6: Contour plots for the first two corrected EOF vectors: (a) – 1$^{\text{st}}$ EOF; and (b) – 2$^{\text{nd}}$ EOF. (White=-1.7; Black=1.4.)

main that is of interest and n, the total number of time points. Henceforth, Bayesian EOFs are proposed in this section as an alternative to classical EOFs.

Bayesian EOFs: the underlying model

The Bayesian EOFs model we adopt here is given by

$$\mathbf{Z} = \mu \otimes \mathbf{1}'_n + \mathbf{\Phi X} + \epsilon, \qquad \epsilon \sim N_{p \times n}(\mathbf{0}, \mathbf{\Sigma}_S \otimes \mathbf{\Sigma}_\epsilon) \qquad (6.6)$$

and

$$\mathbf{X} \,|\theta \sim N_{p \times n}(\mathbf{0}, \mathbf{I}_p \otimes \mathbf{\Sigma}_T), \qquad (6.7)$$

where we assume both spatial and temporal covariance matrices $\mathbf{\Sigma}_S$ and $\mathbf{\Sigma}_T$ have full rank, that is, nonsingular. Moreover, we deal with two cases

138

Index of EOFs	True	Classical	Corrected
1	40.000	40.833	39.507
2	20.000	20.230	20.178
3	15.000	15.890	15.072
4	10.000	9.165	10.267
5	0.023	0.544	0.189
6	0.024	0.519	0.183
7	0.024	0.481	0.178
8	0.024	0.434	0.178
9	0.024	0.431	0.173
10	0.024	0.405	0.171

Table 6.1: Percentage of spatial variation (%) for the first 10 EOFs by the true, classical, and corrected methods ($\rho = 0.9$).

	Matrix discrepancies
Classical v.s. True	4.347
Corrected v.s. True	0.137

Table 6.2: Matrix discrepancies for the classical and corrected EOFs against the true EOFs.

for the temporal covariance matrix $\boldsymbol{\Sigma}_T$: (i) It is randomly distributed with an inverted Wishart distribution; and (ii) It has a semi–parametric form, that is, $\boldsymbol{\Sigma}_T = \sigma^2 \rho(.,\theta)$, with known temporal correlation form for $\rho(.,.)$ but unknown parameters σ^2 and θ. We assume that the temporal correlation matrix, $\rho(.,\theta)$, decreases as the difference between two time points increases, such that $\rho(t_i - t_j, \theta) \rightarrow 1$ as $t_i - t_j \rightarrow 0$. In practice, we assume $\sigma^2 = 1$ due to the non–identifiable property for the Kronecker product. The semi–parametric form for $\rho(.,\theta)$ is then estimable using the MCMC method. For simplicity, we assume no small–scale spatial variation and measurement errors, that is, no ϵ term in (6.6), deferring such refinements to future work.

In (6.6), $\boldsymbol{\Phi}$ can be treated as a constant matrix for a known $\boldsymbol{\Sigma}_S$ by the Karhunen–Loeve (KL) expansion. Given the positive definite spatial covariance function $\boldsymbol{\Sigma}_S : p \times p$, we have the unitary orthogonal matrix $\mathbf{O} : p \times p$, with its first row being positive, and diagonal matrix $\boldsymbol{\Lambda} = \text{diag} \{\lambda_1, \ldots, \lambda_p\} : p \times p$, with λ^2s being the eigenvalues for $\boldsymbol{\Sigma}_S$ and $\lambda_1 > \ldots > \lambda_p > 0$, such

that

$$\boldsymbol{\Sigma}_S^{-1} = \mathbf{O}\boldsymbol{\Lambda}^{-2}\mathbf{O}'.$$

Hence, we have the Karhunen–Loeve (KL) expansion as follows:

$$
\begin{aligned}
\boldsymbol{\Sigma}_S^{-1} &= (\lambda_1^{-1}\mathbf{O}^1,\ldots,\lambda_p^{-1}\mathbf{O}^p)(\lambda_1^{-1}\mathbf{O}^1,\ldots,\lambda_p^{-1}\mathbf{O}^p)' \\
&= (\boldsymbol{\Phi}^1,\ldots,\boldsymbol{\Phi}^p)(\boldsymbol{\Phi}^1,\ldots,\boldsymbol{\Phi}^p)' \\
&= \sum_{j=1}^p \boldsymbol{\Phi}^j\boldsymbol{\Phi}^{j'} \\
&= \boldsymbol{\Phi}\boldsymbol{\Phi}',
\end{aligned}
$$

where $\boldsymbol{\Phi}_j = \lambda_j^{-1}\mathbf{O}^j$, for $j = 1,\ldots,p$, and $\boldsymbol{\Phi} = (\boldsymbol{\Phi}^1,\ldots,\boldsymbol{\Phi}^p) : p \times p$.

However, when $\boldsymbol{\Sigma}_S$ is a random matrix, the EOFs represented by the columns of $\boldsymbol{\Phi}$ are also random. Moreover, the orthogonal matrix can be treated as either constant or random in the KL expansion. The distribution for the random orthogonal matrix has been obtained by James (1954a), as an invariant uniform distribution on the Stiefel manifold (see Definition 6.4.2). Moreover, he also obtained the independent distribution of the diagonal entries in $\boldsymbol{\Lambda}^2$ in the KL expansion.

Then a Bayesian version of EOFs can be obtained either by the MCMC method or through the empirical Bayes approach. The first level of a hierarchical model places no restriction on the form of our Bayesian EOFs, and so it is a nonparametric approach. When the prior for the purely spatial covariance matrix, $\boldsymbol{\Sigma}_S$, has been determined, $\boldsymbol{\Phi}$ can be obtained using the KL expansion and Lemma 6.4.1 below.

Therefore, the Bayesian EOF model (6.6)–(6.7) is completed by specifying prior distributions for the model parameters: $p(\mu)$, $p(\boldsymbol{\Sigma}_T)$ for Case (i) [or $p(\theta)$ for Case (ii)] and $p(\boldsymbol{\Sigma}_S)$. Lacking specific prior information, we assume $p(\mu) \propto 1$ and $\boldsymbol{\Sigma}_S^{-1} \sim W_p(\delta_S, \boldsymbol{\Xi}_S)$ with δ_S and $\boldsymbol{\Xi}_S$ being hyperparameters.

For Case (i), we assume $\boldsymbol{\Sigma}_T^{-1} \sim W_n(\delta_T, \boldsymbol{\Xi}_T)$. The collection of hyperparameters can be denoted by $\mathcal{H}_1 = \{\delta_S, \delta_T, \boldsymbol{\Xi}_S, \boldsymbol{\Xi}_T\}$. The joint posterior distribution we are interested in is given by $p(\mu, \boldsymbol{\Sigma}_T, \boldsymbol{\Sigma}_S|\mathbf{Y})$.

For Case (ii), we assume $\theta \sim N_k(\theta_0, \boldsymbol{\Sigma}_0)$. The collection of hyperparameters here can be denoted by $\mathcal{H}_2 = \{\delta_S, \boldsymbol{\Xi}_S, \theta_0, \boldsymbol{\Sigma}_0\}$. The joint posterior distribution we need for inference is then given by $p(\mu, \theta, \boldsymbol{\Sigma}_S | \mathbf{Y})$.

In summary, we consider the following two Bayesian models to obtain the Bayesian EOFs for each of the two cases under consideration in which we assume no measurement error (ϵs):

(i) The Bayesian model is given by

$$
\begin{aligned}
\mathbf{Z} &= \mu \otimes \mathbf{1}_n' + \boldsymbol{\Phi}\mathbf{X} & (6.8) \\
\mathbf{X} &\sim N_{p \times n}(\mathbf{0}, \mathbf{I}_p \otimes \boldsymbol{\Sigma}_T) & (6.9) \\
\boldsymbol{\Phi} &= \mathbf{O}\boldsymbol{\Lambda}^{-1}, & (6.10)
\end{aligned}
$$

where $\boldsymbol{\Sigma}_S^{-1} = \mathbf{O}\boldsymbol{\Lambda}^{-2}\mathbf{O}'$ by the above KL expansion. And the priors for the model parameters are assumed mutually independent and given as follows:

$$
\begin{aligned}
p(\mu) &\propto 1 & (6.11) \\
\boldsymbol{\Sigma}_S^{-1} &\sim W_p(\delta_S, \boldsymbol{\Xi}_S) & (6.12) \\
\boldsymbol{\Sigma}_T^{-1} &\sim W_n(\delta_T, \boldsymbol{\Xi}_T). & (6.13)
\end{aligned}
$$

(ii) The Bayesian model is given as in (6.8)–(6.10), but $\boldsymbol{\Sigma}_T$ can be written as $\rho(., \theta)$, where $\rho(., .)$ is the known temporal correlation matrix, decreasing as the difference between any two time points increases, and θ is an unknown parameter. The priors for μ and $\boldsymbol{\Sigma}_S^{-1}$ are given in (6.11)–(6.12), respectively, and they are mutually independent. Model specification is completed with the prior for the temporal covariance function given by:

$$
\theta \sim N_k(\theta_0, \boldsymbol{\Sigma}_0). \qquad (6.14)
$$

Notice that we assume both $\boldsymbol{\Sigma}_S$ and $\boldsymbol{\Sigma}_T$ are valid covariance matrices and non–singular, that is, the rank for $\boldsymbol{\Sigma}_S$ is p and for $\boldsymbol{\Sigma}_T$, n.

Decomposition of $\mathbf{Y} \sim N_{p \times n}(\mathbf{0}, \boldsymbol{\Sigma}_S \otimes \mathbf{I}_n)$

We first describe results from James (1954a), who discovers that the distribution of the mean $\mathbf{0}$, temporal independent random matrix $\mathbf{Y}$ [that is, $\mathbf{Y} \sim N_{p \times n}(\mathbf{0}, \boldsymbol{\Sigma}_S \otimes \mathbf{I}_n)$] can be uniquely decomposed into three independent parts, specifically, one part being Wishart distributed, one part being uniformly distributed on a Grassmann manifold and the last part being uniformly distributed on a Stiefel manifold, that is, an orthogonal group in this setting. He constructs invariant measures on the orthogonal group (that is, the Haar measure), the Grassmann and Stiefel manifolds. James also finds the distribution of a non–central Wishart distribution using the Haar measure (1954b) and the distribution of the latent roots for a covariance matrix (1960). We now introduce some basic notation from James (1954a) and Chikuse (2004).

Definition 6.4.1 *The orthogonal group, $O(n)$, is the set of all orthogonal matrices with the operation of matrix multiplication.*

Definition 6.4.2 *The Stiefel manifold, $V_{k,n} = \{ \boldsymbol{V} : n \times k;\ \boldsymbol{V}' \boldsymbol{V} = \boldsymbol{I}_k \}$, is a set of k $(k \leq n)$ orthonormal matrices in $\mathcal{R}^n$.*

Definition 6.4.3 *The Grassmann manifold, $G_{k,n-k}$, is the set of all k–dimensional hyperplanes in $\mathcal{R}^n$ that pass through the origin.*

James sates that "…the Grassmann and Stiefel manifolds may be regarded as coset spaces of the orthogonal group" (1954, p.63), an important property on their relationships. The main result from James has been summarized in the following lemma.

Lemma 6.4.1 *(James, 1945) Suppose $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) \sim N_{p \times n}(\boldsymbol{0}, \boldsymbol{\Sigma}_S \otimes \boldsymbol{I}_n)$. Then we have*

$$\boldsymbol{Y} \;=\; \boldsymbol{O}\boldsymbol{L}\boldsymbol{P}, \tag{6.15}$$

where $\boldsymbol{O} : p \times p$ represents an orthogonal matrix that is uniformly distributed over the Grassmann manifold, $\boldsymbol{P} : p \times n$, a semi–orthogonal matrix (i.e.,

142

$\boldsymbol{PP'} = \boldsymbol{I}_p$ *in this case) that is uniformly distributed over the Stiefel manifold and* $\boldsymbol{L} : p \times p$, *a diagonal matrix with diagonal entries* $\{l_1, \ldots, l_p\}$ *such that* $l_1^2, \ldots, l_p^2$ *are the eigenvalues for* $\boldsymbol{YY'}$ *with* $l_1 > \ldots > l_p > 0$.

Mardia and Khatri (1977) develop the exact and asymptotic distributions for the random matrix uniformly distributed on a Stiefel manifold. They also discuss the matrix form of the von Mises–Fisher distribution on a Stiefel manifold. We will not discuss this application of James in this chapter but leave the construction of the posterior distributions on Stiefel and Grassmann manifolds for future research.

Theoretical results

We present the related inference for Bayesian EOFs in following lemmas and theorems. All proofs in this section are listed in Appendix C.1.

Lemma 6.4.1 and the KL expansion below provide the basis for the theoretical results needed to obtain Bayesian EOFs. It leads to the prior distribution as a special case when $\boldsymbol{\Xi}_S = \boldsymbol{I}_p$, as shown in following lemma.

Lemma 6.4.2 *If* $\boldsymbol{\Sigma}_S^{-1} \sim W_p(n, \boldsymbol{I}_p)$ *for some* $n \in \mathcal{Z}^+$, *by the KL expansion and Lemma 6.4.1, we have that the* λ^{-2}'s *are mutually independently distributed with* $\lambda_j^{-2} \sim \chi_n^2$, *for* $j = 1, \ldots, p$, *and* $\boldsymbol{O} : p \times p$ *is uniformly distributed on the Stiefel manifold.*

Lemma 6.4.2 provides one way to sample the random matrix $\boldsymbol{\Sigma}_S^{-1}$ from its prior distribution $W_p(\delta_S, \boldsymbol{I}_p)$.

Theorem 6.4.1 *Consider the data matrix* $\boldsymbol{Y} : p \times n \sim N_{p \times n}(\boldsymbol{0}, \boldsymbol{\Sigma}_S \otimes \boldsymbol{\Sigma}_T)$. *Given the nonsingular spatial and temporal covariance matrices* $\boldsymbol{\Sigma}_S$ *and* $\boldsymbol{\Sigma}_T$, *let* $\boldsymbol{Y}^* = \boldsymbol{Y}\boldsymbol{\Sigma}_T^{-1/2}$. *Then* $\boldsymbol{Y}^* \sim N_{p \times n}(\boldsymbol{0}, \boldsymbol{\Sigma}_S \otimes \boldsymbol{I}_n)$. *Consequently,* $\boldsymbol{Y}^* = \boldsymbol{\Sigma}_S^{1/2}\boldsymbol{OLP}$, *where* $\boldsymbol{O}$, $\boldsymbol{L}$, *and* $\boldsymbol{P}$ *are given in Lemma 6.4.1. The Bayesian EOFs are then obtained as* $\boldsymbol{W} = \frac{1}{n}\boldsymbol{\Sigma}_S^{1/2}\boldsymbol{OL}$.

Theorem 6.4.2 *Consider the data matrix* $\boldsymbol{Y} : p \times n \sim N_{p \times n}(\boldsymbol{0}, \boldsymbol{\Sigma}_S \otimes \boldsymbol{\Sigma}_T)$. *Suppose the temporal covariance matrix* $\boldsymbol{\Sigma}_T$ *is nonsingular and known. Then*

we have $\boldsymbol{Y}^* = \boldsymbol{Y}\boldsymbol{\Sigma}_T^{-1/2} \sim N_{p \times n}(\boldsymbol{0}, \boldsymbol{\Sigma}_S \otimes \boldsymbol{I}_n)$. *Assume* $\boldsymbol{\Sigma}_S^{-1} \sim W_p(\delta_S, \boldsymbol{\Xi}_S)$. *The posterior distribution for the spatial precision matrix* $\boldsymbol{\Sigma}_S^{-1}$ *is given as follows:*

$$\boldsymbol{\Sigma}_S^{-1}|\boldsymbol{Y} \quad \sim \quad W_p(\delta^o, \boldsymbol{\Xi}^o), \tag{6.16}$$

where $\delta^o = \delta_S + n$, *and*

$$\boldsymbol{\Xi}^o \quad = \quad \boldsymbol{\Xi}_S - \boldsymbol{\Xi}_S\,\boldsymbol{Y}(\boldsymbol{Y}'\boldsymbol{\Xi}_S\,\boldsymbol{Y} + \boldsymbol{\Sigma}_T)^{-1}\,\boldsymbol{Y}'\boldsymbol{\Xi}_S. \tag{6.17}$$

The Bayesian EOFs can be obtained when $\boldsymbol{\Sigma}_S$ *is estimated or sampled from its posterior distribution.*

Consequently from Theorem 6.4.2, we can either obtain the estimates for $\boldsymbol{\Sigma}_S$ using an empirical method such as that of the Sampson–Guttorp (SG) or sample it from its posterior distribution in (6.16). In other words, we can either use empirical Bayes or hierarchical Bayesian methods to obtain the estimates for the model parameters.

If the spatial covariance matrix were known, valid and nonsingular, we would have a similar result for the posterior distribution of the temporal covariance matrix, $\boldsymbol{\Sigma}_T$. The next theorem tells us its posterior distribution if the prior for $\boldsymbol{\Sigma}_T^{-1}$ is assumed to be Wishart distributed, that is, Case (i).

Theorem 6.4.3 *Consider the data matrix* $\boldsymbol{Y} : p \times n \sim N_{p \times n}(\boldsymbol{0}, \boldsymbol{\Sigma}_S \otimes \boldsymbol{\Sigma}_T)$. *Suppose the spatial covariance matrix* $\boldsymbol{\Sigma}_S$ *is known, valid and nonsingular. Assume the prior for* $\boldsymbol{\Sigma}_T^{-1}$ *is* $W_n(\delta_T, \boldsymbol{\Xi}_T)$. *Then the posterior distribution for* $\boldsymbol{\Sigma}_T^{-1}$ *is given by*

$$\boldsymbol{\Sigma}_T^{-1}|\boldsymbol{Y} \quad \sim \quad W_n(\delta^*, \boldsymbol{\Xi}^*), \tag{6.18}$$

where $\delta^* = \delta_T + p$ *and*

$$\boldsymbol{\Xi}^* \quad = \quad \boldsymbol{\Xi}_T - \boldsymbol{\Xi}_T\,\boldsymbol{Y}'(\boldsymbol{Y}\boldsymbol{\Xi}_T\,\boldsymbol{Y}' + \boldsymbol{\Sigma}_S)^{-1}\,\boldsymbol{Y}\boldsymbol{\Xi}_T. \tag{6.19}$$

In Case (ii) where the temporal covariance matrix is assumed to have a

known parametric form but unknown parameters, we can obtain the posterior distribution for these parameters given a valid and known spatial covariance matrix. This posterior distribution is given in the following theorem.

Theorem 6.4.4 *Given the same condition as in Theorem 6.4.3 and the fully Bayesian model in Case (ii) with the priors for θ in (6.14), the conditional posterior distributions for θ are given as follows:*

$$
p(\theta|\boldsymbol{Y}) \quad \propto \quad \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma^2}tr(\boldsymbol{V}'\boldsymbol{V}\rho(.,\theta)^{-1}) + (\theta - \theta_0)'\boldsymbol{\Sigma}_0^{-1}(\theta - \theta_0)\right]\right\},
$$

(6.20)

where $\boldsymbol{V} = \boldsymbol{\Sigma}_S^{-1/2}\boldsymbol{Y}$.

In practice, both temporal and spatial covariance matrices are unknown, and so conditions in Theorems 6.4.1 and 6.4.4 do not hold. Although the nonsingularity condition for these two covariance matrices might be difficult to verify due to the challenging numerical problems, we need the condition to obtain the posterior samples for both matrices in this chapter. Future research will be devoted to improve on the results.

To obtain posterior samples for both parameters in the MCMC framework, we can either use a mixture of MCMC and empirical Bayes methods or use pure MCMC runs. We will illustrate the algorithm for both cases next. But the idea here is to obtain the conditional posterior samples for μ, $\boldsymbol{\Sigma}_S$ and $\boldsymbol{\Sigma}_T$ subsequently. The empirical Bayes method will be used to obtain the estimate for $\boldsymbol{\Sigma}_S$ given the data matrix, μ, and $\boldsymbol{\Sigma}_T$. We use the Gelman and Rubin R statistics as a device to check the convergence of the Markov chains (Gelman et al., 2004, p. 296–297). The estimates for the model parameters are then obtained as the mean of the posterior samples after the burn–in period.

We next develop the posterior conditional distributions for the model parameters for Cases (i) and (ii). The Bayesian EOFs can be obtained by Theorem 6.4.1 in which the mean field, spatial and temporal covariance

matrices are estimated from their posterior distributions or by the empirical Bayes method.

Posterior conditional distributions

The posterior conditional distributions for μ, $\boldsymbol{\Sigma}_S^{-1}$, and $\boldsymbol{\Sigma}_T^{-1}$ can be obtained for Cases (i) and (ii) in the fully Bayesian framework. The proofs of the theorems in this section are presented in Appendix C.1. We first present the conditional posterior distributions for model parameters for Case (i) in the following theorem.

Theorem 6.4.5 *Given the Bayesian hierarchical model in (6.8)–(6.13), the posterior conditional distributions for these model parameters are given as follows:*

(i) *The conditional posterior distribution for μ is given by*

$$\mu | \boldsymbol{Z}, \boldsymbol{\Sigma}_S, \boldsymbol{\Sigma}_T \quad \sim \quad N_p(\boldsymbol{M}, \Sigma^* \boldsymbol{\Sigma}_S) \tag{6.21}$$

where $\Sigma^* = \{ tr(\boldsymbol{1}_n' \boldsymbol{1}_n \boldsymbol{\Sigma}_T^{-1}) \}^{-1}$ and $\boldsymbol{M} = \boldsymbol{Z} \boldsymbol{\Sigma}_T^{-1} \boldsymbol{1}_n' \Sigma^*$.

(ii) *The conditional posterior distribution for $\boldsymbol{\Sigma}_S^{-1}$ is given by*

$$\boldsymbol{\Sigma}_S^{-1} | \boldsymbol{Z}, \mu, \boldsymbol{\Sigma}_T \quad \sim \quad W_p(\delta_1, \boldsymbol{\Xi}_1), \tag{6.22}$$

where $\delta_1 = \delta_S + n$,

$$\boldsymbol{\Xi}_1 = \boldsymbol{\Xi}_s - \boldsymbol{\Xi}_s \boldsymbol{Y} (\boldsymbol{Y}' \boldsymbol{\Xi}_S \boldsymbol{Y} + \boldsymbol{\Sigma}_T)^{-1} \boldsymbol{Y}' \boldsymbol{\Xi}_s, \tag{6.23}$$

and $\boldsymbol{Y} = \boldsymbol{Z} - \mu \otimes \boldsymbol{1}_n'$.

(iii) *The conditional posterior distribution for $\boldsymbol{\Sigma}_T^{-1}$ is given by*

$$\boldsymbol{\Sigma}_T^{-1} | \boldsymbol{Z}, \mu, \boldsymbol{\Sigma}_S \quad \sim \quad W_n(\delta_2, \boldsymbol{\Xi}_2), \tag{6.24}$$

where $\delta_2 = \delta_T + p$,

$$\boldsymbol{\Xi}_2 = \boldsymbol{\Xi}_T - \boldsymbol{\Xi}_T \boldsymbol{Y}' (\boldsymbol{Y} \boldsymbol{\Xi}_T \boldsymbol{Y}' + \boldsymbol{\Sigma}_S) \boldsymbol{Y} \boldsymbol{\Xi}_T, \tag{6.25}$$

146

and $\boldsymbol{Y} = \boldsymbol{Z} - \mu \otimes \boldsymbol{1}'_n$.

In the same way, the posterior conditional distributions for the model parameters for Case (ii) are obtained in the next theorem.

Theorem 6.4.6 *Given the Bayesian hierarchical model in (6.8)–(6.12) and (6.14), the posterior conditional distributions for μ and Σ_S^{-1} are given in Theorem 6.4.5. Let $\boldsymbol{V} = \boldsymbol{\Sigma}_S^{-1/2}(\boldsymbol{Z} - \mu \otimes \boldsymbol{1}'_n)$. Moreover, the posterior conditional distribution for θ is given by:*

$$p(\theta|\boldsymbol{Z}, \mu, \boldsymbol{\Sigma}_S) \quad \propto \quad \exp\left\{ -\frac{1}{2}\left[tr\left(\boldsymbol{V}\boldsymbol{V}'\rho(.,\theta)^{-1} \right) + (\theta - \theta_0)'\boldsymbol{\Sigma}_0^{-1}(\theta - \theta_0) \right] \right\}.$$

$$(6.26)$$

After obtaining the estimates for the spatial and temporal covariance matrices, as well as other model parameters for both cases, the Bayesian EOFs can then be obtained by Theorem 6.4.1, Lemma 6.4.1 and the KL expansion.

Next we illustrate how the MCMC algorithm can be used to obtain the posterior samples from the joint posterior distribution $p(\mu, \boldsymbol{\Sigma}_S^{-1}, \boldsymbol{\Sigma}_T^{-1}|\mathbf{Z})$ and $p(\mu, \boldsymbol{\Sigma}_S^{-1}, \theta|\mathbf{Z})$ for Cases (i) and (ii), respectively.

MCMC algorithms

To obtain the posterior samples from the joint posterior distribution of model parameters, $\mu, \boldsymbol{\Sigma}_S$ and $\boldsymbol{\Sigma}_T$ (or θ), the MCMC algorithm is used to draw samples from their posterior distributions. For Case (i), we use Gibbs sampling method based on the full conditional distributions we obtained before. For Case (ii), a Metropolis–within–Gibbs algorithm is used because the posterior conditional distribution for θ does not have any closed form.

Algorithm 6.4.1 *For Case (i), Gibbs sampling can be used to draw the posterior samples from $p(\mu, \boldsymbol{\Sigma}_S, \boldsymbol{\Sigma}_T|\boldsymbol{Z})$:*

1. Initialization: set

$$\mu^{(1)} = \overline{\boldsymbol{Z}}^{row},$$

sample

$$\mathbf{\Sigma}_S^{-1^{(1)}} \sim W_p(\delta_S, \mathbf{\Xi}_S),$$

and

$$\mathbf{\Sigma}_T^{-1^{(1)}} \sim W_n(\delta_T, \mathbf{\Xi}_T).$$

2. Given the $(j-1)^{th}$ values, $\mu^{(j-1)}$, $\mathbf{\Sigma}_S^{-1^{(j-1)}}$, $\mathbf{\Sigma}_T^{-1^{(j-1)}}$, and $\mathbf{Z}$:

(1) Sample $\mu^{(j)}$ from $p(\mu|\mathbf{Z}, \mathbf{\Sigma}_S^{(j-1)}, \mathbf{\Sigma}_T^{(j-1)})$ from (6.21).

(2) Sample $\mathbf{\Sigma}_S^{-1^{(j)}}$ from $p(\mathbf{\Sigma}_S^{-1}|\mathbf{Z}, \mu^{(j)}, \mathbf{\Sigma}_T^{(j-1)})$ from (6.22).

(3) Sample $\mathbf{\Sigma}_T^{-1^{(j)}}$ from $p(\mathbf{\Sigma}_T^{-1}|\mathbf{Z}, \mu^{(j)}, \mathbf{\Sigma}_S^{(j)})$ from (6.24).

3. Repeat until convergence.

The Metropolis–within–Gibbs algorithm is omitted here because we present a similar result in Chapter 2. In this section, we give the algorithm for a very special case when the temporal process is assumed to be an AR(1) process. Hence we have ϕ as the parameter that characterizes the AR(1) process. Assume that $\phi \sim N(\phi_0, \sigma_{\phi_0}^2)$. Then the collection of hyperparameters can be denoted by $\mathcal{H} = \{\phi_0, \sigma_{\phi_0}^2, \delta_S, \mathbf{\Xi}_S\}$. Since there is no closed form for the posterior conditional distribution in (6.26), the Metropolis–within–Gibbs algorithm could then be used to draw posterior samples of interest.

The next section includes a straightforward extension on the Bayesian EOFs results where its spatial covariance is assumed to have a GIW distribution instead of IW (see Le and Zidek (2006), for example).

6.5 Extension to the Bayesian EOFs

We can extend the above results about Bayesian EOFs to incorporate the GIW prior for the spatial covariance structure in such a way that the GIW prior reflects some characteristics of the data matrix. Le & Zidek (1992–2006) develop theoretical results for modelling spatio–temporal processes, i.e., the BSP approach in Chapter 4.

In such a Gaussian GIW framework, the spatial covariance matrix $\mathbf{\Sigma}_S$ can be estimated by the SG–method or Damian SG–method (Damian et al., 2002). Therefore, the estimates for the $\mathbf{\Sigma}_S$ can be updated at each iteration in the MCMC sampling. This will be carried out in future work.

The next section includes two simulation examples that help assess the performance for the Bayesian, classical and corrected EOFs.

6.6   Simulation Study 2

The Bayesian EOF models we consider above have a very general structure. Note that ϵ in (6.6) represents small scale spatial variation or measurement error. If ϵ is close to $\mathbf{0}$, or equivalently to a very small value for $\mathbf{\Sigma}_\epsilon$, we then have approximately $\mathbf{Y} = \mu \otimes \mathbf{1}'_n + \mathbf{\Phi X}$, where $\mathbf{\Phi}$'s columns are the EOFs for $\mathbf{\Sigma}_S$ and $\mathbf{X}|\theta \sim N_{p \times n}(\mathbf{0}, \mathbf{I}_p \otimes \mathbf{\Sigma}_T(\theta))$. If $\mathbf{\Sigma}_T(\theta) = \mathbf{I}_n$, then we have the classical EOFs in Section 6.2. If $\mathbf{\Sigma}_T(\theta) \neq \mathbf{I}_n$ but known, we then have the corrected EOFs in Section 6.3. If $\mathbf{\Sigma}_T(\theta)$ is unknown, we can use the Bayesian EOFs obtained in Section 6.4.

The objective in this example is to compare the three different types of EOFs for a separable state–space process with known spatial and temporal covariance matrices. To do that, we first simulate the matrix–variate data set. We then compute these three EOFs and compare them with the true EOFs by contour plots and the matrix discrepancies.

To briefly review these three types of EOFs, suppose $\mathbf{Y} : p \times n$ represents the anomaly matrix for p sites and n time points, and follows a multivariate normal distribution $N_{p \times n}(\mathbf{0}, \mathbf{\Sigma}_S \otimes \mathbf{\Sigma}_T)$. Recall that the classical EOFs estimate the sample spatial covariance matrix by $\frac{1}{n}\mathbf{Y}\mathbf{Y}'$. Given the temporal dependence structure, that is, the temporal covariance matrix, the corrected EOFs estimate the sample spatial covariance matrix by $\mathbf{Y}\mathbf{\Sigma}_T^{-1}\mathbf{Y}'$. Given the priors for the spatial and temporal covariance matrix, the Bayesian EOFs estimate the sample spatial covariance matrix in the hierarchical model by means of the corresponding posterior mode.

We consider two cases in this section to assess the performance of the EOFs for two different temporal dependence structures. For both cases, we

assume a separable space time covariance structure, that is, an exponential spatial covariance function and an AR(1) temporal covariance function. In particular, the spatial covariance function is given by

$$(\mathbf{\Sigma}_S)_{ij} = \exp(-V_{ij}/\lambda), \tag{6.27}$$

where V_{ij} is the Euclidean distance between $\mathbf{s}_i$ and $\mathbf{s}_j$, for $i, j = 1, \ldots, p$ and λ, a scale parameter. The temporal covariance function between t_k and t_l is given by $\sigma_v^2 \phi^{2\|t_k - t_l\|}$, for $t_k, t_l \in \{1, \ldots, T\}$. Note that $|\phi| < 1$ indicates a causal AR(1) process. If $\phi \simeq 0$, then $\mathbf{y}_t$ are approximately independent; if $\phi \simeq 1$, then $\{\mathbf{y}_t : t = 1, \ldots, n\}$ is a highly autocorrelated AR(1) process. We consider Case (i), $\phi = 0.1$, and Case (ii), $\phi = 0.9$.

The geographical region in this simulation study is set to be $[0.1, 1.0] \times [0.1, 1.0]$. We select 100 grid points in this region to be the locations of interest, i.e., $p = 100$. We then choose $n = 120$ time points at each of these 100 sites.

The initial settings for the separable space–time covariance functions are given as follows: $\lambda = 0.4$, $\sigma_v^2 = 1.0$, and $\phi = 0.1$ for Case (i), and 0.9 for Case (ii).

We define $Y(\mathbf{s}, t) = Z(\mathbf{s}, t) - \hat{\mu}(\mathbf{s})$, the anomaly at site $\mathbf{s}$ and time t, where $\hat{\mu}(\mathbf{s}) = \frac{1}{n} \sum_{t=1}^{n} Z(\mathbf{s}, t)$. We obtain the classical, corrected, and Bayesian EOFs using $Y(\mathbf{s}, t)$.

We now compare the corrected, classical and "true" EOFs in the these two cases: (i) $\phi = 0.1$, and (ii) $\phi = 0.9$, respectively.

Simulated data

Suppose $\mathbf{Y} : p \times n \sim N_{p \times n}(\mathbf{0}, \mathbf{\Sigma}_S \otimes \mathbf{\Sigma}_T)$. One way to generate the simulated data is by first simulating $\mathbf{Y}^* = \mathbf{Y}\mathbf{\Sigma}_T^{1/2} = (\mathbf{y}_1^*, \ldots, \mathbf{y}_n^*)$. Thus, $\mathbf{y}_t^* \sim N_p(\mathbf{0}, \mathbf{\Sigma}_S)$, independently for $t = 1, \ldots, n$. We then generate $\mathbf{Y}$ by $\mathbf{Y}^* \mathbf{\Sigma}_T^{-1/2}$.

An alternative to obtain the simulated data uses James's result and the KL expansion. Given both $\mathbf{\Sigma}_S$ and $\mathbf{\Sigma}_T$, we first illustrate the way to generate the simulated data in any given regions. Given the spatial covariance matrix

$\mathbf{\Sigma}_S$, the Karhunen–Loeve expansion gives the unique orthogonal matrix $\mathbf{O}$: $p{\times}p$ with its first row being positive and the unique diagonal matrix $\mathbf{\Lambda}^2 : p{\times}p$ with its decreasing diagonal entries being the eigenvalues of $\mathbf{\Sigma}_S$, such that

$$
\begin{aligned}
\mathbf{\Sigma}_S &= \mathbf{O}\mathbf{\Lambda}^2\mathbf{O}' \\
&= (\mathbf{\Lambda}\mathbf{O}')'\mathbf{\Lambda}\mathbf{O}' \\
&= (\lambda_1\mathbf{O}^{(1)},\ldots,\lambda_p\mathbf{O}^{(p)})(\lambda_1\mathbf{O}^{(1)},\ldots,\lambda_p\mathbf{O}^{(p)})' \\
&= (\mathbf{\Phi}^{(1)},\ldots,\mathbf{\Phi}^{(p)})(\mathbf{\Phi}^{(1)},\ldots,\mathbf{\Phi}^{(p)})' \\
&= \sum_{j=1}^{p}\mathbf{\Phi}^{(j)}(\mathbf{\Phi}^{(j)})' \\
&= \mathbf{\Phi}\mathbf{\Phi}',
\end{aligned}
\tag{6.28}
$$

where $\mathbf{\Phi} = (\mathbf{\Phi}^{(1)},\ldots,\mathbf{\Phi}^{(p)}) : p \times p$ and $\mathbf{\Phi}^{(j)} = \lambda_j\mathbf{O}^{(j)}$ for $j = 1,\ldots,p$.

Given $\mathbf{\Sigma}_T : n \times n$, the corresponding Karhunen–Loeve expansion is given by

$$
\begin{aligned}
\mathbf{\Sigma}_T &= \mathbf{P}\mathbf{L}^2\mathbf{P}' \\
&= \sum_{j=1}^{p}\mathbf{\Psi}^{(j)}(\mathbf{\Psi}^{(j)})' \\
&= \mathbf{\Psi}\mathbf{\Psi}',
\end{aligned}
\tag{6.29}
$$

where $\mathbf{P} : n \times n$ represents the orthogonal matrix with its first row positive; $\mathbf{L}^2$ is a diagonal matrix with decreasing but positive diagonal entries $l_1^2,\ldots,l_n^2$; $\mathbf{\Psi} = (\mathbf{\Psi}^{(1)},\ldots,\mathbf{\Psi}^{(n)})$, and $\mathbf{\Psi}^{(i)} = l_i\mathbf{P}^{(i)}$ for $i = 1,\ldots,n$.

Consequently, if $\mathbf{Y} : p{\times}n \sim N_{p\times n}(\mathbf{0}, \mathbf{\Sigma}_S{\otimes}\mathbf{\Sigma}_T)$, and $\mathbf{\Xi} \sim N_{p\times n}(\mathbf{0}, \mathbf{I}_p{\otimes}\mathbf{I}_n)$, then $\mathbf{Y} = \mathbf{\Phi}\mathbf{\Xi}\mathbf{\Psi}'$ where $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are given by (6.28) and (6.29), respectively. Let $\mathbf{\Xi}^* = \mathbf{\Xi}\mathbf{\Psi}'$. Then $\mathbf{\Xi}^* \sim N_{p\times n}(\mathbf{0}, \mathbf{I}_p \otimes \mathbf{\Sigma}_T)$, and $\mathbf{Y} = \mathbf{\Phi}\mathbf{\Xi}^*$.

In short, the simulated data can be generated, for known spatial and temporal covariance matrices, as follows:

(i) Uniquely obtain $\mathbf{\Phi}$ and $\mathbf{\Psi}$ as in (6.28) and (6.29) for known $\mathbf{\Sigma}_S$ and $\mathbf{\Sigma}_T$, respectively.

(ii) Generate $\mathbf{\Xi} \sim N_{p\times n}(\mathbf{0}, \mathbf{I}_p{\otimes}\mathbf{I}_n)$ by n samples independently distributed

from the multivariate normal distribution $N_p(\mathbf{0}, \mathbf{I}_p)$.

(iii) Obtain the simulated data matrix $\mathbf{Y} : p \times n$ by $\mathbf{\Phi \Xi \Psi'}$.

Results for Case (i): $\rho = 0.1$

Table 6.3 demonstrates the percentage of the spatial variance of each spatial pattern found by the EOFs against the total space variance. It shows that these percentages for the true, classical or corrected EOFs are quite close to each other. Table 6.4 presents the matrix discrepancies of the corrected or classical EOFs to that of the true EOFs. These two matrix discrepancies are close to each other but the classical one is slightly better than the corrected one.

Index of EOFs	True	Classical	Corrected
1	33.584	31.913	31.883
2	11.122	13.476	13.457
3	11.122	8.963	8.996
4	5.108	6.193	6.201
5	3.853	4.732	4.747
6	3.638	3.467	3.467
7	2.274	2.733	2.735
8	2.274	2.333	2.329
9	1.537	1.946	1.945
10	1.537	1.719	1.724

Table 6.3: Percentage of spatial variation (%) for the first 10 EOFs by the true, classical, and corrected methods ($\rho = 0.1$).

	Matrix discrepancies
Classical v.s. True	0.230
Corrected v.s. True	0.233

Table 6.4: Matrix discrepancies for the classical and corrected EOFs against the true EOFs ($\rho = 0.1$).

Figures 6.7–6.9 plot the contours for the three types of EOFs. In Figure 6.7, the first EOF in (a) shows an ellipsoidal spatial pattern in this field,

with its mode at the center of this region. The second EOF in (b) shows a north–east to south–west shifting spatial pattern with the negative mode at the north–east corner and the positive one at the south–west corner. The third EOF in (c) shows an approximately opposite spatial pattern to that in (b). The remaining three EOFs in (d)–(f) also represent certain spatial patterns in this field. It shows the classical EOFs to be quite similar to the corrected one due to the very small value of ϕ. Moreover, the ratio of matrix discrepancy between the classical and true EOFs matrix is 0.25, while that between the corrected and true ones, 0.25. It verifies our expectation that both EOFs work quite well since the "true" data are approximately independent samples across the overall locations. Classical EOFs are supposed to capture the main types of spatial patterns in this case.
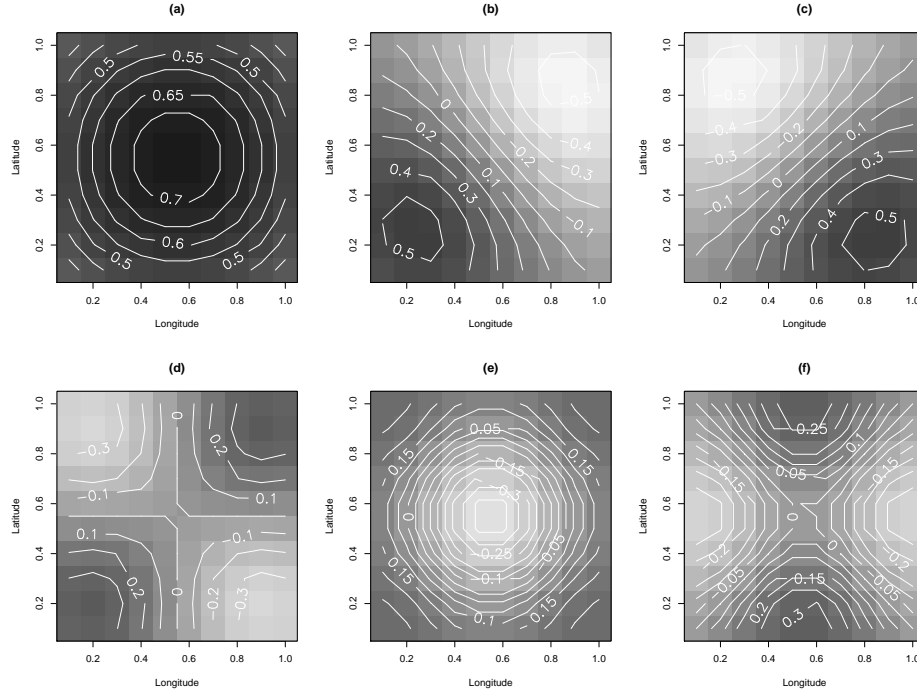
Figure 6.7: Contour plots for the first 6 true EOFs ($\rho = 0.1$): (a) – 1st EOF; (b) – 2nd EOF; (c) – 3rd EOF; (d) – 4th EOF; (e) – 5th EOF; and (f) – 6th EOF. (White=-0.6; Black=0.9.)
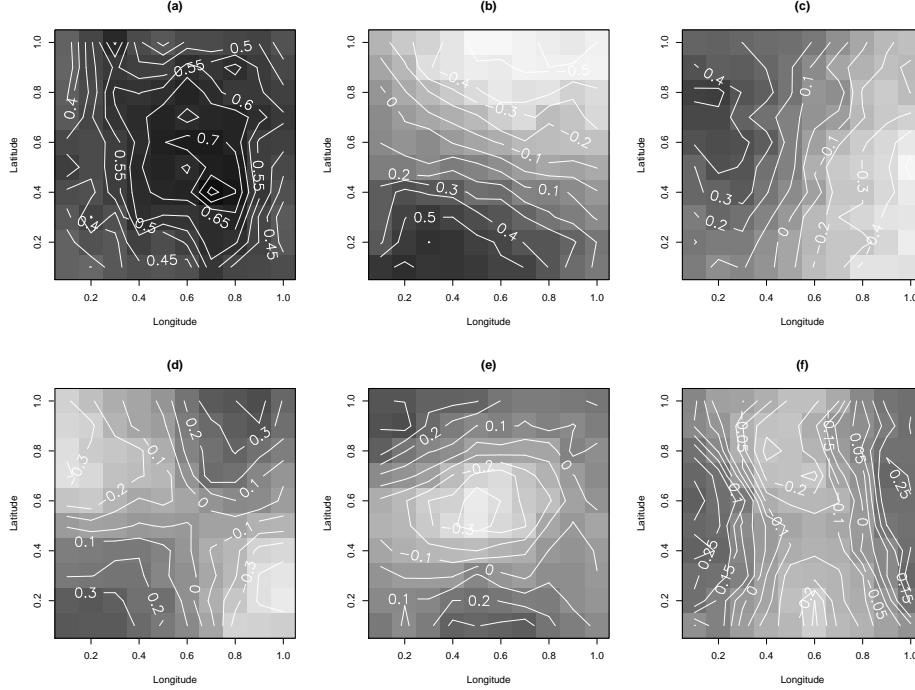
Figure 6.8: Contour plots for the first 6 classical EOFs ($\rho = 0.1$): (a) – 1st EOF; (b) – 2nd EOF; (c) – 3rd EOF; (d) – 4th EOF; (e) – 5th EOF; and (f) – 6th EOF. (White=-0.6; Black=0.9.)

Results for Case (ii): $\rho = 0.9$

Table 6.5 plots the percentages of the spatial variations for each of the three types EOFs: true, classical, and corrected. This graph shows that the corrected EOFs gives more accurate estimates on the majority diagonal elements of $\mathbf{\Lambda}$ in the KL expansion. Table 6.6 presents the matrix discrepancies between the corrected and classical EOFs and the true EOFs. It shows that the matrix discrepancy of the corrected EOFs against the true ones is much smaller than that of the classical EOFs against the true ones, which implies more accurate results of the corrected EOFs than those of the classical ones.

Figures 6.10–6.11 present the first six classical and corrected EOFs for this case. Comparing them with Figure 6.7, it is obviously that the corrected EOFs can estimate the main types of spatial patterns better than

154
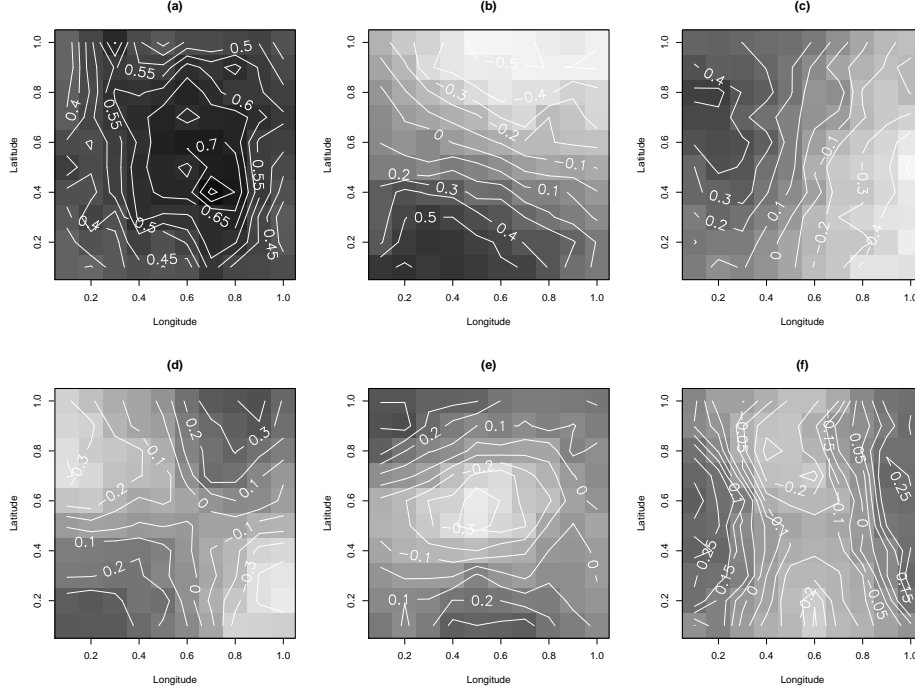
Figure 6.9: Contour plots for the first 6 corrected EOFs ($\rho = 0.1$): (a) – 1st EOF; (b) – 2nd EOF; (c) – 3rd EOF; (d) – 4th EOF; (e) – 5th EOF; and (f) – 6th EOF. (White=-0.6; Black=0.9.)

the classical ones. Moreover, the ratio of matrix discrepancy between the classical and true EOFs matrix is 5.79, while that between the corrected and true ones, 0.21. It shows that the classical EOFs are far from the "truth" for the highly autocorrelated database.

6.7   Conclusions

We have proposed Bayesian EOF (BEOF) approach in this chapter and show the corresponding theoretical results as well as the MCMC algorithm to obtain the posterior samples of the model parameters. We have shown that the corrected EOF method can be used to obtain better representation of principal spatial patterns than the classical EOF method in two simula-

Index of EOFs	True	Classical	Corrected
1	33.584	33.572	37.702
2	11.122	20.789	11.868
3	11.122	9.660	11.005
4	5.108	7.835	6.204
5	3.853	4.341	4.344
6	3.638	3.302	3.405
7	2.274	2.870	2.038
8	2.274	2.287	1.910
9	1.537	1.746	1.521
10	1.537	1.583	1.460

Table 6.5: Percentage of spatial variation (%) for the first 10 EOFs by the true, classical, and corrected methods ($\rho = 0.9$).

	Matrix discrepancies
Classical v.s. True	4.978
Corrected v.s. True	0.332

Table 6.6: Matrix discrepancies for the classical and corrected EOFs against the true EOFs ($\rho = 0.9$).

tion studies. From where, we conclude that the classical EOF may lead to severe problems for a highly temporally correlated space–time process. The corrected EOF greatly improves the performance of the classical EOF and captures the principal spatial patterns better than the classical one. The implementation of the BEOF method will be one future work, as well as the comparisons among the BEOFs, corrected and classical EOFs.

Here notice that the BEOF is different than the Bayesian factor analysis proposed by Aguilar and West (2000). In the Bayesian factor analysis they defined, PCA is used for the time–varying Σ_S. And MCMC algorithm has to be used to draw posterior samples for the Bayesian factors, which is computationally costly. We so extend the BEOF into an extension of the BSP to model the univariate or multivariate responses in spatio–temporal fields, which we call generalized Bayesian spatial prediction (GBSP) method.
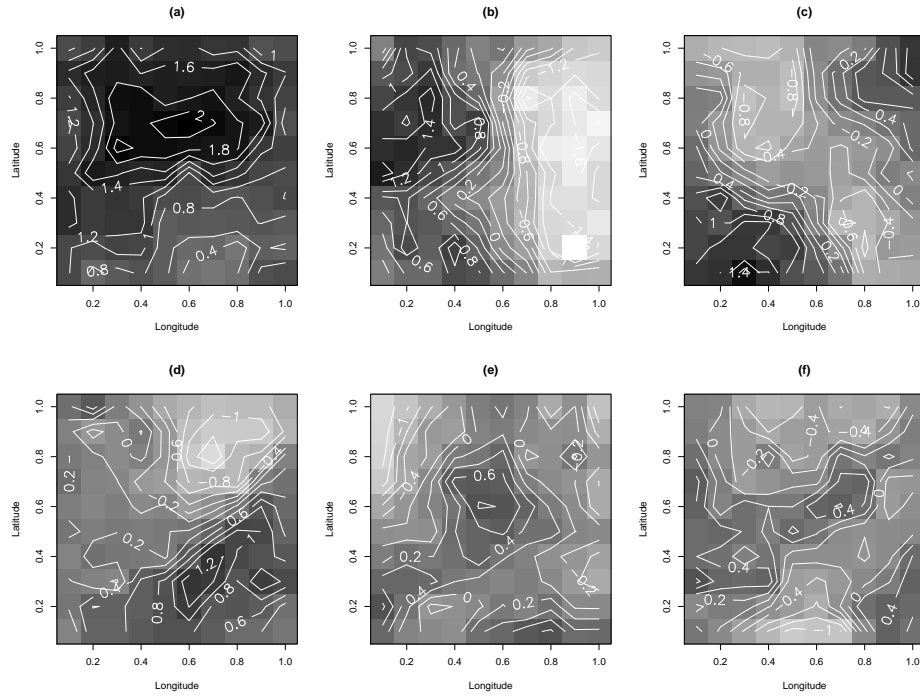
Figure 6.10: Contour plots for the first 6 classical EOFs ($\rho = 0.9$): (a) – 1$^{\text{st}}$ EOF; (b) – 2$^{\text{nd}}$ EOF; (c) – 3$^{\text{rd}}$ EOF; (d) – 4$^{\text{th}}$ EOF; (e) – 5$^{\text{th}}$ EOF; and (f) – 6$^{\text{th}}$ EOF. (White=-1.6; Black=2.2.)
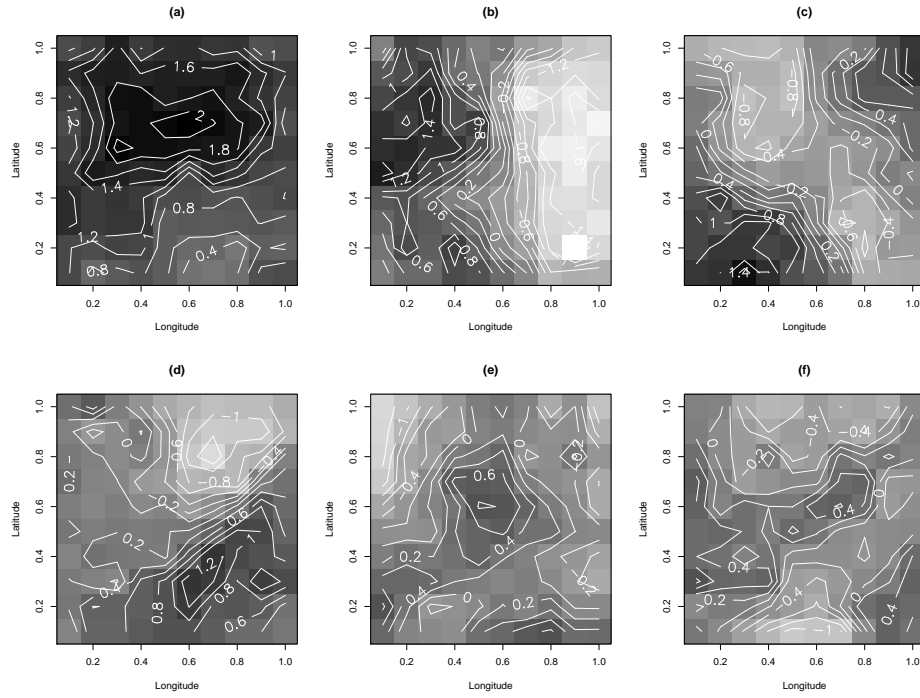
Figure 6.11: Contour plots for the first 6 corrected EOFs ($\rho = 0.9$): (a) – 1st EOF; (b) – 2nd EOF; (c) – 3rd EOF; (d) – 4th EOF; (e) – 5th EOF; and (f) – 6th EOF. (White=-1.6; Black=2.2.)

Chapter 7

An Extension of the BSP: Bayesian Spatio–Temporal Models

This chapter proposes an extension of the BSP approach for modelling air pollution data in large spatial–temporal domains, such as the AQS database we discussed in Chapters 2 and 3. The motivation for this extension has been addressed in previous studies. We extend the BSP approach because of its computational efficiency and better model performance in spatial interpolation and temporal prediction than those of the DLM. We integrate the BSP into the DLM framework because the DLM has a fairly flexible structure for temporal prediction at ungauged and gauged locations and capability for handling the missing observations at gauged sites.

Our proposed model can deal with two types of covariates in this field: time–varying but site invariant covariates and site–specific and time–varying covariates. We decompose the underlying space–time processes into three parts: a long–term spatial–temporal term, a short–term main spatial pattern and a short–term spatial–temporal term. We also incorporate Bayesian EOFs into the new model so we can model gridded data, such as the MAQSIP (Multiscale Air Quality Simulation Platform) simulated data. We also extend the univariate Bayesian spatio–temporal model to the multivariate case. Moreover, we summarize the MCMC method that allows us to draw MC samples from the joint posterior distribution of model parameters. And henceforth, we are able to interpolate or predict the univariate (or multivariate) pollutant(s) in large spatio–temporal domains.

Section 7.1 introduces related research in this field. Section 7.2 proposes a univariate Bayesian spatio–temporal model, an extension of the BSP approach, to model spatio–temporal processes over large domains. Section 7.3 discusses the relationships between our model and some others, such as the DLM proposed by Huerta et al. (2004), the state–space model (SSM) by Wikle and Cressie (1999), Gelfand et al. (2005), and Le–Zidek's approach (1992). Section 7.4 presents the multivariate Bayesian spatio–temporal model. Section 7.5 demonstrates use of the MCMC algorithm to draw samples from the joint posterior distribution of the model parameters. Section 7.6 summarizes the advantages for the Bayesian spatio–temporal models we propose in this chapter and the remaining future work for our models.

7.1 Introduction

In large spatial–temporal domains, computational efficiency often emerges as a major difficulty in modelling high–dimensional data. This computational problem has been addressed by many approaches.

Mardia et al. (1998) propose the kriged Kalman filter (KKF) approach in which the mean structure in the observation equation are formed by finite common fields. However, in the discussion following Mardia et al.'s paper, Cressie and Wikle (1998) criticize the KKF approach for its oversmoothing predictor and lack of specific prior structure for some of the model parameters, comparing with the SSM proposed by Wikle and Cressie (1999), a fully hierarchical Bayesian approach. They use EOFs based on the orthogonal basis functions to capture the main spatial patterns in spatial–temporal fields and so tackle the problem of the curse of dimensionality. However, Wikle and Cressie's SSM can only incorporate an AR(1) process, and so would be inapplicable to more general structures of autocorrelated residuals after removing the long–term spatial–temporal component from the observation model.

Gelfand et al. (2005) develop a dynamic spatio–temporal model using the linear coregionalization method (LMC) to incorporate different correlation

structure for multivariate responses. However, their model is not applicable to autocorrelated detrended residuals. A similar approach can also be seen in Lee and Ghosh (2005). Johannesson et al. (2007) propose dynamic multi–resolution spatial models (MRSM) to extend the MRSM and incorporate dynamic temporal features. Moreover, the spatial domain considered in their approach is recursively partitioned with a "tree–structured parent–children relationship between cells (pixels) at adjacent resolution". Lopes et al. (2007) propose a spatial dynamic factor analysis to model the common factors as a latent process with unknown finite dimension, much smaller than the total number of sites. The forward–filtering–backward–sampling method is then applied to the latent process to obtain the posterior samples for state parameters in the latent process.

Other novel works related to geostatistics can be referred to Wacker-nagel (1998). A fully Bayesian approach on kriging has been investigated by Handcock and Stein (1993). They state the ordinary kriging can be viewed as the Bayesian kriging under a non–informative prior for the mean. In their approach, they incorporate the unknown covariance structure in a Bayesian framework. The Bayesian kriging approach takes into account more uncertainty on the unknown covariance structure and so is capable of quantifying the performance of the estimated kriging predictor.

Le and Zidek developed their approach in accord with a specific set of desiderata, one of which was computational feasibility. Moreover their applications were commonly made to relatively compact regions such as urban areas. These two factors led them to use conventional EDA and time series approaches to initially remove regional temporal components by fitting identical model parameters to all sites, making the standard errors of the resulting estimates negligible. In one case where there method was employed to cover a very large spatial domain (Fu et al., 2003) a spatial thin plate spline was fitted over sites so, in effect, a different mean level $\mu(\mathbf{s})$ was subtracted from each site's series so as to center it. Thus in most cases, Le and Zidek focus on modeling small scale variation. At the same time, even though the large scale components were handled in a fairly ad hoc way, the approach itself was very general and very flexible. Thus for example, it sub-

sumes the approach of Wikle and Cressie (1999), albeit without the elegant modeling foundations the latter provide and the substantive rationale they give for it.

Our approach attempts to meet their desideratum of computational simplicity while putting their prefiltering approach into a more formal hierarchical framework. That retains the advantages of their methodology. At the same time, it extends the approach of Wikle and Cressie (1999) and includes a number of other approaches for modelling space time processes. Thus it provides a unifying Bayesian framework wherein computational strategies can be developed once and for all. That in turn gives modellers much flexibility.

Next we propose the extension of the BSP approach, that is, the univariate and multivariate Bayesian spatio–temporal models.

7.2 Univariate Bayesian Spatio–temporal Model

We consider the case that the observations are measured at irregular locations in a large spatial domain. We propose this univariate Bayesian spatio–temporal model to spatially interpolate and temporally predict the response variables at those irregularly located gauged or ungauged sites. One future implementation will be to predict ground–level ozone concentrations over the eastern USA using the hourly ozone AQS database in Section 2.5. Comparing it with Wikle and Cressie's approach, our model decomposes the spatial–temporal process into the following components: a long–term dynamic spatio–temporal term, a short–term spatial patterns and a short–term spatial temporal component. Note that Wikle and Cressie (1999) assume that long–term spatial–temporal components are the averages over all locations, and so they consider the response to be the detrended residuals.

Denote by $Z(\mathbf{s}, t)$ the observation at site $\mathbf{s}$ and time t. We model the spatio–temporal fields by first removing all the linear and seasonal trends across all regions in the domain that is of interest, that is, $\mathbf{F}'_1(\mathbf{s}, t)\mathbf{M}(\mathbf{s}, t)$. The remaining detrended residuals are represented by $Y(\mathbf{s}, t)$. We then decompose $Y(\mathbf{s}, t)$ into two parts, $\mathbf{\Phi}^K(\mathbf{s})\mathbf{a}^K(t)$ – the remaining long–term spa-

162

tial variation, and $V(\mathbf{s}, t)$ – the short–term spatial–temporal variation. In our approach, this short–term spatial–temporal term can be modelled as a BSP term in Le–Zidek's approach, incorporated into a DLM framework.

We propose the following model in this section:

$$Z(\mathbf{s}, t) = \mathbf{F}'_1(\mathbf{s}, t)\mathbf{M}(\mathbf{s}, t) + Y(\mathbf{s}, t) \qquad (7.1)$$

$$Y(\mathbf{s}, t) = \mathbf{\Phi}^K(\mathbf{s})a^K(t) + V(\mathbf{s}, t) \qquad (7.2)$$

$$V(\mathbf{s}, t) = \mathbf{F}'_2(\mathbf{s}, t)\theta(\mathbf{s}, t) + \nu(\mathbf{s}, t) \qquad (7.3)$$

and

$$\mathbf{M}(\mathbf{s}, t) = \mathbf{G}_1(\mathbf{s}, t)\mathbf{M}(\mathbf{s}, t-1) + \eta(\mathbf{s}, t) \qquad (7.4)$$

$$\mathbf{a}^K(t) = \mathbf{H}_{2t}\mathbf{a}^K(t-1) + \xi_t \qquad (7.5)$$

$$\theta(\mathbf{s}, t) = \mathbf{G}_2(\mathbf{s}, t)\theta(\mathbf{s}, t-1) + \omega(\mathbf{s}, t). \qquad (7.6)$$

The above models can incorporate a fairly large class of cases, such as systematic temporal components, regional level covariate effects, and local level or site–specific covariate effects. In the coming subsections, we discuss in particular the way to deal with different kinds of covariates in our model and possible choices for the terms to illustrate the covariance structure at the coarse level, that is, $\mathbf{\Phi}^K(\mathbf{s})$. Specifically, we consider two types of covariates: Type I covariates are the regional level covariates, common over all sites at any fixed time point, with a site–specific coefficient matrix; Type II covariates are the local level covariates, site–specific and time–varying, with a common coefficient vector across all sites at any fixed time point.

The vector form of the above proposed univariate Bayesian spatio–temporal model can be written as follows:

$$\mathbf{Z}_t = \mathbf{F}'_{1t}\mathbf{M}_t + \mathbf{Y}_t \qquad (7.7)$$

$$\mathbf{Y}_t = \mathbf{\Phi}^K\mathbf{a}^K(t) + \mathbf{V}_t \qquad (7.8)$$

$$\mathbf{V}_t = \mathbf{F}'_{2t}\theta_t + \nu_t \qquad (7.9)$$

$$\mathbf{M}_t = \mathbf{G}_{1t}\mathbf{M}_{t-1} + \eta_t \tag{7.10}$$

$$\mathbf{a}^K(t) = \mathbf{H}_{2t}\mathbf{a}^K(t-1) + \xi_t \tag{7.11}$$

$$\theta_t = \mathbf{G}_{2t}\theta_{t-1} + \omega_t, \tag{7.12}$$

where $\mathbf{Z}_t = (Z(\mathbf{s}_1, t), \ldots, Z(\mathbf{s}_p, t))' : p \times 1$, $\mathbf{F}'_{1t} = \mathrm{diag}\{F_1(\mathbf{s}_1, t), \ldots, F_1(\mathbf{s}_p, t)\}$, $\mathbf{M}_t = (M(\mathbf{s}_1, t), \ldots, M(\mathbf{s}_p, t))'$, $\mathbf{Y}_t = (Y(\mathbf{s}_1, t), \ldots, Y(\mathbf{s}_p, t))'$, $\mathbf{\Phi}^K = (\Phi^K(s_1)$, $\ldots, \Phi^K(s_p))'$, $\mathbf{V}_t = (V(\mathbf{s}_1, t), \ldots, V(\mathbf{s}_p, t))'$, $\mathbf{F}'_{2t} = \mathrm{diag}\{F_2(\mathbf{s}_1, t), \ldots,$ $F_2(\mathbf{s}_p, t)\}$, $\theta_t = (\theta(\mathbf{s}_1, t), \ldots, \theta(\mathbf{s}_p, t))'$, $\nu_t = (\nu(\mathbf{s}_1, t), \ldots, \nu(\mathbf{s}_p, t))'$, $\eta_t = (\eta(\mathbf{s}_1, t), \ldots, \eta(\mathbf{s}_p, t))'$, $\xi_t = (\xi(\mathbf{s}_1, t), \ldots, \xi(\mathbf{s}_p, t))'$, $\omega_t = (\omega(\mathbf{s}_1, t), \ldots, \omega(\mathbf{s}_p, t))'$ and $G_{jt} = \mathrm{diag}\{G_j(\mathbf{s}_1, t), \ldots, G_j(\mathbf{s}_p, t)\}$ for $j = 1, 2$.

Our proposed univariate Bayesian spatio–temporal model is completed by giving the following assumptions on the priors of model parameters:

$$\mathbf{\Sigma}^C = \mathbf{O}_C \mathbf{\Lambda}_C^{-2} \mathbf{O}'_C \tag{7.13}$$

$$\mathbf{\Phi} = \mathbf{O}_C \mathbf{\Lambda}_C^{-1} \tag{7.14}$$

$$\mathbf{\Phi}^K = \mathbf{\Phi} \mathbf{E}_p^K \tag{7.15}$$

$$\nu_t \sim N_p(\mathbf{0}, \mathbf{\Sigma}^F) \tag{7.16}$$

$$\eta_t \sim N_{l_1}(\mathbf{0}, \mathbf{W}_1) \tag{7.17}$$

$$\xi_t \sim N_K(\mathbf{0}, \mathbf{W}^K) \tag{7.18}$$

$$\omega_t \sim N_{l_2}(\mathbf{0}, \mathbf{W}_2) \tag{7.19}$$

$$\mathbf{\Sigma}^{-C} \sim W_p(\mathbf{\Xi}^C, \delta^C) \tag{7.20}$$

$$\mathbf{\Sigma}^{-F} \sim W_p(\mathbf{\Xi}^F, \delta^F), \tag{7.21}$$

where $\mathbf{\Sigma}^C$ presents the covariance matrix for $Y(\mathbf{s}, t)$ at the coarse level and henceforth $\mathbf{\Sigma}^{-C}$, the precision matrix; similarly, $\mathbf{\Sigma}^F$ represents the covariance matrix at the fine level and $\mathbf{\Sigma}^{-F}$, the precision matrix. Let $\mathbf{E}_p^K : p \times K = (\mathbf{e}_{p,1}, \ldots, \mathbf{e}_{p,K})$, where $\mathbf{e}_{p,j} : p \times 1$ represents the p–dimensional vector whose j^{th} entry is 1, the others, 0, for $j = 1, \ldots, K$.

The initial information can be described as follows:

$$\mathbf{M}_0 \sim N(\mathbf{m}_0^M, \mathbf{C}_0^M) \tag{7.22}$$

$$\mathbf{a}^K(0) \sim N(\mathbf{m}_0^a, \mathbf{C}_0^a) \tag{7.23}$$

$$\theta_0 \sim N(\mathbf{m}_0^\theta, \mathbf{C}_0^\theta). \tag{7.24}$$

In the coming subsections, we first discuss possible choices and models for different types of covariates, along with the main spatial term contributed by Φ^K, and then obtain theoretical results for our proposed model.

7.2.1 Type I covariates

We can incorporate Type I covariates in our model, that is, the covariates are common over sites at each fixed time point, for example, the month effect, week day effect, and hourly effect we considered in Chapter 4. For these types of covariate, we consider their coefficient matrix (unknown parameter matrix) to be site–specific. In this subsection, we show how we deal with this type of covariate as the first step in (7.1) and (7.4).

Suppose the data base has l_1 Type I covariates in total. Let $\breve{\mathbf{Z}}_t = (\breve{Z}_1(t), \ldots, \breve{Z}_{l_1}(t)) : 1 \times l_1$ be the l_1–dimension covariate vector at time t. The corresponding coefficient matrix $\breve{\mathbf{B}} : l_1 \times p$ can also be written as $\breve{\mathbf{B}} = (\breve{\beta}(\mathbf{s}_1), \ldots, \breve{\beta}(\mathbf{s}_p))'$, where $\breve{\beta}(\mathbf{s}_i) = (\breve{\beta}_1(\mathbf{s}_i), \ldots, \breve{\beta}_{l_1}(\mathbf{s}_i))$ is the l_1–dimensional coefficient column vector at site $\mathbf{s}_i$ for $i = 1, \ldots, p$.

Let $\mathbf{F}_1'(\mathbf{s}_i, t) = \breve{\mathbf{Z}}_t$ and $\mathbf{M}(\mathbf{s}_i, t) = \breve{\beta}(\mathbf{s}_i)$. Then (7.1) is equivalent to $Z(\mathbf{s}_i, t) = \breve{\mathbf{Z}}_t \breve{\beta}(\mathbf{s}_i) + Y(\mathbf{s}_i, t)$. This shows our model to incorporate Type I covariates.

7.2.2 Type II covariates

We now show how we can deal with Type II covariates in this subsection using the model we proposed above, that is, site specific covariates varying with time, for example, hourly temperatures, wind speeds and wind directions at each location. The corresponding coefficients of Type II covariates are common over all sites at any fixed time point. We show this type of covariates can be dealt with at steps (7.3) and (7.6), along with an AR(1) autocorrelation structure for the detrended residuals $V(\mathbf{s}_i, t)$. Of course, the case of other systematic temporal components can also be dealt with using

a similar approach.

Suppose l_2 Type II covariates all are of interest. Let $\tilde{\mathbf{Z}}_t(\mathbf{s}_i) = (\tilde{Z}_{1,t}(\mathbf{s}_i), \ldots,$ $\tilde{Z}_{l_2,t}(\mathbf{s}_i))$ be the l_2–dimensional covariate vector at time t and location $\mathbf{s}_i$, for $i = 1, \ldots, p$ and $t = 1, \ldots, n$. Then its corresponding coefficients are constant across all sites at each fixed time point t, that is, $\tilde{\beta}'_t = (\tilde{\beta}_{1,t}, \ldots, \tilde{\beta}_{l_2,t}) : 1 \times l_2$.

Let $\mathbf{F}'_1(\mathbf{s}_i, t) = \tilde{\mathbf{Z}}_t(\mathbf{s}_i)$ and $\mathbf{M}(\mathbf{s}_i, t) = \tilde{\beta}_t$. Then (7.1) is equivalent to $Z(\mathbf{s}_i, t) = \tilde{\mathbf{Z}}_t(\mathbf{s}_i)\tilde{\beta}_t + Y(\mathbf{s}_i, t)$, demonstrating our model's flexibility.

7.2.3 Possible choices for $\mathbf{\Phi}^K(s)$

$\mathbf{\Phi}^K(\mathbf{s})$ can be chosen in various ways. For example, it could be formed from polynomials. It can also be formed from the empirical orthogonal functions as discussed in Chapter 6. We illustrate some possible choices for $\mathbf{\Phi}^K(\mathbf{s})$ in this subsection.

The simplest case would be the polynomials. For example, $\mathbf{\Phi}^K(\mathbf{s})$ can come from a polynomial in latitude and longitude, representing a linear mean surface of the spatial–temporal field of $Y(\mathbf{s}, t)$ in the geographical coordinates (Stroud et al., 2001). Of course, $\mathbf{\Phi}^K(\mathbf{s})$ can be constructed using other polynomial basis functions, such as the quartic polynomials, that is, fourth degree polynomials, used by Fuentes and Raftery (2005).

Or $\mathbf{\Phi}^K(\mathbf{s})$ could also be orthogonal basis functions. For example, it could be orthogonal polynomials, orthonormal basis functions for cubic splines, bicubic splines, wavelets, or empirical orthogonal functions. For the reasons we stated before in Chapter 6, we emphasize EOF functions. Wikle and Cressie (1999) estimate the EOF in their SSM, which could do here. For unknown EOFs, we extend the Bayesian EOFs in Chapter 6 under the univariate Bayesian spatio–temporal model. We will talk about this in the coming subsection.

7.2.4 Predictive posterior distributions

We present theoretic results for the predictive posterior distribution of model parameters used in our model. Moreover, we briefly summarize the idea of implementing the BEOF in our model and the related theoretical results.

Suppose $\mathbf{\Sigma}^C$ denotes the coarse–level covariance matrix for $Y(\mathbf{s}, t)$. Assume $\mathbf{\Sigma}^F$ is the fine–level covariance matrix for the deAR'd residuals, included in the covariance matrix of $\nu(\mathbf{s}, t)$.

The univariate Bayesian spatio–temporal model we proposed in (7.7)–(7.12) can also be written as the following form:

$$
\begin{aligned}
\mathbf{Z}_t &= \mathbf{F}'_t \theta^*_t + \nu_t & \nu_t &\sim N_p(\mathbf{0}, \mathbf{\Sigma}^{\mathrm{F}}) \\
\theta^*_t &= \mathbf{G}_t \theta^*_{t-1} + \omega^*_t & \omega^*_t &\sim N_{l_1+K+l_2}(\mathbf{0}, \mathbf{W}),
\end{aligned}
$$

where $\mathbf{F}'_t = (\mathbf{F}'_{1t}, \mathbf{\Phi}^K, \mathbf{F}'_{2t})$, $\theta^*_t = (\mathbf{M}'_t, \mathbf{a}^K(t)', \theta'_t)'$, $\mathbf{G}_t = \text{Block–diag}\{\mathbf{G}_{1t}, \mathbf{H}_{2t}, \mathbf{G}_{2t}\}$, $\omega^*_t = (\eta'_t, \xi'_t, \omega'_t)'$, and $\mathbf{W} = \text{Block–diag}\{\mathbf{W}_1, \mathbf{W}^K, \mathbf{W}_2\}$.

Using standard results for the DLM and referring to Theorem A.2.1 in Appendix A.2, we obtain the corresponding posterior distributions for the state parameters, that is, θ^*_t, given the observations until time t and the coarse– and fine– level covariance matrices $\mathbf{\Sigma}^C$ and $\mathbf{\Sigma}^F$, respectively, in the following theorem.

Theorem 7.2.1 *Given the coarse– and fine– level covariance matrices, $\mathbf{\Sigma}^C$ and $\mathbf{\Sigma}^F$, respectively, we obtain the following posterior distributions:*

$$
\begin{aligned}
\theta^*_{t-1} | \mathbf{Z}_{1:t-1}, \mathbf{\Sigma}^F, \mathbf{\Sigma}^C &\sim N_{l_1+K+l_2}[\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}] \\
\theta^*_t | \mathbf{Z}_{1:t-1}, \mathbf{\Sigma}^F, \mathbf{\Sigma}^C &\sim N_{l_1+K+l_2}[\boldsymbol{a}_{t-1}, \boldsymbol{R}_{t-1}] \\
\mathbf{Z}_t | \mathbf{Z}_{1:t-1}, \mathbf{\Sigma}^F, \mathbf{\Sigma}^C &\sim N_p[\boldsymbol{f}_t, \boldsymbol{Q}_t] \\
\theta^*_t | \mathbf{Z}_{1:t}, \mathbf{\Sigma}^F, \mathbf{\Sigma}^C &\sim N_{l_1+K+l_2}[\boldsymbol{m}_t, \boldsymbol{C}_t],
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{a}_t &= \mathbf{G}_t \boldsymbol{m}_{t-1} & \boldsymbol{R}_t &= \mathbf{G}_t \boldsymbol{C}_{t-1} \mathbf{G}'_t + \mathbf{W} \\
\boldsymbol{f}_t &= \mathbf{F}'_t \boldsymbol{a}_t & \boldsymbol{Q}_t &= \mathbf{F}'_t \boldsymbol{R}_t \mathbf{F}_t + \mathbf{\Sigma}^F \\
\boldsymbol{e}_t &= \mathbf{Z}_t - \boldsymbol{f}_t & \boldsymbol{A}_t &= \boldsymbol{R}_t \mathbf{F}_t \boldsymbol{Q}_t^{-1} \\
\boldsymbol{m}_t &= \boldsymbol{a}_t + \boldsymbol{A}_t \boldsymbol{e}_t & \boldsymbol{C}_t &= \boldsymbol{R}_t - \boldsymbol{A}_t \boldsymbol{Q}_t \boldsymbol{A}'_t,
\end{aligned}
$$

for $t = 1, \ldots, n$.

Interest lies in the joint posterior distribution $p(\theta^*_{1:n}, \mathbf{\Sigma}^C, \mathbf{\Sigma}^F | \mathbf{Z}_{1:n})$. We can write this joint posterior density as the product of $p(\theta^*_{1:n} | \mathbf{\Sigma}^C, \mathbf{\Sigma}^F, \mathbf{Z}_{1:n})$ and $p(\mathbf{\Sigma}^C, \mathbf{\Sigma}^F | \mathbf{Z}_{1:n})$. By Theorem 7.2.1, the former posterior distribution can

be obtained as follows:

$$
\begin{aligned}
p(\theta_{1:n}^*|\mathbf{\Sigma}^{\mathrm{C}}, \mathbf{\Sigma}^{\mathrm{F}}, \mathbf{Z}_{1:n}) &= \prod_{t=1}^{n} p(\theta_t^*|\mathbf{\Sigma}^{\mathrm{C}}, \mathbf{\Sigma}^{\mathrm{F}}, \mathbf{Z}_{1:t}) p(\theta_0^*|\mathbf{\Sigma}^{\mathrm{C}}, \mathbf{\Sigma}^{\mathrm{F}}) \\
&\propto \prod_{t=1}^{n} |\mathbf{Q}_t|^{-1/2} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{n} (\theta_t^* - \mathbf{m}_t)' \mathbf{C}_t^{-1} (\theta_t^* - \mathbf{m}_t) \right\}.
\end{aligned}
$$

The latter is given by

$$
\begin{aligned}
p(\mathbf{\Sigma}^{\mathrm{C}}, \mathbf{\Sigma}^{\mathrm{F}}|\mathbf{Z}_{1:n}) &\propto p(\mathbf{Z}_{1:n}|\mathbf{\Sigma}^{\mathrm{C}}, \mathbf{\Sigma}^{\mathrm{F}}) p(\mathbf{\Sigma}^{\mathrm{C}}) p(\mathbf{\Sigma}^{\mathrm{F}}) \\
&= \prod_{t=1}^{n} p(\mathbf{Z}_t|\mathbf{\Sigma}^{\mathrm{C}}, \mathbf{\Sigma}^{\mathrm{F}}, \mathbf{Z}_{1:t-1}) p(\mathbf{\Sigma}^{\mathrm{C}}) p(\mathbf{\Sigma}^{\mathrm{F}}) \\
&\propto \left(\prod_{t=1}^{n} |\mathbf{Q}_t| \right)^{-1/2} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{n} \mathbf{e}_t' \mathbf{Q}_t^{-1} \mathbf{e}_t \right\} p(\mathbf{\Sigma}^{\mathrm{C}}) p(\mathbf{\Sigma}^{\mathrm{F}}),
\end{aligned}
$$

where $\mathbf{e}_t$ and $\mathbf{Q}_t$ are given in Theorem 7.2.1. Note that $\mathbf{F}_t'$ can be viewed as a function of $\mathbf{\Sigma}^{\mathrm{C}}$, that is, $\mathbf{F}_t'(\mathbf{\Sigma}^{\mathrm{C}})$ and so are both $\mathbf{e}_t$ and $\mathbf{Q}_t$.

Moreover, we obtain the following non–analytic form for the full conditional posterior distributions of $\mathbf{\Sigma}^{\mathrm{C}}$ and $\mathbf{\Sigma}^{\mathrm{F}}$:

$$
\begin{aligned}
p(\mathbf{\Sigma}^{\mathrm{F}}|\mathbf{\Sigma}^{\mathrm{C}}, \mathbf{Z}_{1:n}) &\propto p(\mathbf{\Sigma}^{\mathrm{C}}, \mathbf{\Sigma}^{\mathrm{F}}|\mathbf{Z}_{1:n}) \\
&\propto p(\mathbf{\Sigma}^{\mathrm{F}}) \left(\prod_{t=1}^{n} |\mathbf{Q}_t| \right)^{-1/2} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{n} \mathbf{e}_t' \mathbf{Q}_t^{-1} \mathbf{e}_t \right\} \\
&\propto |\mathbf{\Sigma}^{\mathrm{F}}|^{-\frac{\delta_F+p}{2}} \prod_{t=1}^{n} |\mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + \mathbf{\Sigma}^{\mathrm{F}}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \left[\mathrm{tr}(\mathbf{\Xi}^{\mathrm{F}} \mathbf{\Sigma}^{\mathrm{F}-1}) \right. \right. \\
&\qquad \left. \left. + \sum_{t=1}^{n} \mathbf{e}_t' (\mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + \mathbf{\Sigma}^{\mathrm{F}})^{-1} \mathbf{e}_t \right] \right\},
\end{aligned}
\qquad (7.25)
$$

and similarly,

$$
\begin{aligned}
p(\mathbf{\Sigma}^{\mathrm{C}}|\mathbf{\Sigma}^{\mathrm{F}}, \mathbf{Z}_{1:n}) &\propto |\mathbf{\Sigma}^{\mathrm{C}}|^{-\frac{\delta_C+p}{2}} \prod_{t=1}^{n} |\mathbf{F}_t'(\mathbf{\Sigma}^{\mathrm{C}}) \mathbf{R}_t \mathbf{F}_t(\mathbf{\Sigma}^{\mathrm{C}}) + \mathbf{\Sigma}^{\mathrm{F}}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \right. \\
&\qquad \left. \times \left[\mathrm{tr}(\mathbf{\Xi}^{\mathrm{C}} \mathbf{\Sigma}^{\mathrm{C}-1}) + \sum_{t=1}^{n} \mathbf{e}_t'(\mathbf{\Sigma}^{\mathrm{C}}) \left(\mathbf{F}_t'(\mathbf{\Sigma}^{\mathrm{C}}) \mathbf{R}_t \mathbf{F}_t(\mathbf{\Sigma}^{\mathrm{C}}) \right) \right. \right.
\end{aligned}
$$

$$+\mathbf{\Sigma}^{\mathrm{F}})^{-1}\,\mathbf{e}_t(\mathbf{\Sigma}^{\mathrm{C}})\Big]\Big\}\,.$$

$$(7.26)$$

These results give us a theoretical basis for making inferences based on the joint posterior distribution of the model parameters using the MCMC method. The way using the MCMC method will be addressed in Section 7.5.

7.3 The Univariate Bayesian Spatio–temporal Model and Relationships with Others Approaches

We talk about the flexibility of our univariate Bayesian spatio–temporal model and its relationship with some others approaches.

7.3.1 Relationship with the DLM in Huerta et al. (2004)

The long–term spatial temporal term, $\mathbf{F}'_1(\mathbf{s},t)\mathbf{M}(\mathbf{s},t)$, is very general. It contains the case of linear trends and seasonal trends. As a very simple example, the linear trends can be incorporated into our model if we assume $\mathbf{F}'_1(\mathbf{s},t) = (1,t)$ and $\mathbf{M}(\mathbf{s},t)' = (\beta_{0t},\beta_{1t})$. Another example is the DLM in Huerta et al. (2004), a special case of our model. Assume $\mathbf{F}'_1(t) = (1, S_{1t}(a_1), S_{2t}(a_2))$ and $M(\mathbf{s},t)' = (\beta_t, \alpha_{1,s,t}, \alpha_{2,s,t})$. Then we obtain $Z(\mathbf{s},t) = \beta_t + S_{1t}(a_1)\alpha_{1,s,t} + S_{2t}(a_2)\alpha_{2,s,t} + Y(\mathbf{s},t)$, the DLM of Huerta et al. (2004).

7.3.2 Relationship with the SSM in Wikle & Cressie (1999)

Our model contains the special case of the model proposed in Wikle and Cressie (1999), and so can treat the former as an extension of their model. They remove the linear and seasonal trends before fitting their model, a process incorporated by $\mathbf{F}'_1(\mathbf{s},t)\mathbf{M}(\mathbf{s},t)$ in our model. Assuming $V(\mathbf{s},t) = \nu(\mathbf{s},t) + \epsilon(\mathbf{s},t)$, and removing (7.3), (7.4) and (7.6), we obtain the Wikle–Cressie model.

169

Our model can incorporate more general structures for systematic temporal components than Wikle and Cressie's AR(1). To illustrate this generality, we show now an AR(2) process can be dealt with using our model. This would make Wikle and Cressie's model unsuitable for our purpose of modelling hourly ozone concentrations' fields, although it might work for daily ozone concentrations.

Suppose we consider l_1 Type I covariates in (7.3) and AR(2) structure for the detrended residuals. In other words, we consider $X(\mathbf{s}_i, t) = V(\mathbf{s}_i, t) - \mathbf{z}_t\beta(\mathbf{s}_i)$ and $W(\mathbf{s}_i, t) = X(\mathbf{s}_i, t) - \phi_1(\mathbf{s}_i)X(\mathbf{s}_i, t-1) - \phi_2(\mathbf{s}_i)X(\mathbf{s}_i, t-2)$. In (7.3), let $\mathbf{F}'_2(\mathbf{s}_i, t) = (z_t - \phi_1(\mathbf{s}_i)z_{t-1} - \phi_2(\mathbf{s}_i)z_{t-2}, \phi_1(\mathbf{s}_i), \phi_2(\mathbf{s}_i)) : 1 \times (l_1 + 2)$, and $\theta(\mathbf{s}_i, t)' = (\beta(\mathbf{s}_i)', V(\mathbf{s}_i, t-1), V(\mathbf{s}_i, t-2)) : 1 \times (l_1 + 2)$. We then obtain $V(\mathbf{s}_i, t) = F'_2(\mathbf{s}_i, t)\theta(\mathbf{s}_i, t) + W(\mathbf{s}_i, t)$. In (7.6), we further let

$$\mathbf{G}_2(\mathbf{s}_i, t) = \begin{pmatrix} \mathbf{I}_{l_1} & \mathbf{0}_{l_1} & \mathbf{0}_{l_1} \\ \mathbf{z}^*(\mathbf{s}_i, t) & \phi_1(\mathbf{s}_i) & \phi_2(\mathbf{s}_i) \\ \mathbf{0}'_{l_1} & 1 & 0 \end{pmatrix},$$

where $\mathbf{z}^*(\mathbf{s}_i, t) = \mathbf{z}_{t-1} - \phi_1(\mathbf{s}_i)\mathbf{z}_{t-2} - \phi_2(\mathbf{s}_i)\mathbf{z}_{t-3}$. We so obtain (7.6) where $\omega(\mathbf{s}_i, t)' = (\mathbf{0}_{l_1}, W(\mathbf{s}_i, t-1), 0)$. We have shown that our model can deal with AR(2) autocorrelation structure for the detrended residuals. More generally, we can show this dynamic linear modelling approach can actually accommodate far more general time series structure, such as ARMA(p, q) processes. However, Wikle and Cressie only consider the AR(1) process. From that point of view, our approach can be viewed as an extension to their well known model.

7.3.3 Relationship with the univariate SSM in Gelfand et al. (2005)

We now show the relationship between our model and the univariate SSM proposed by Gelfand et al. (2005). They deal with both types of covariates in their model. Suppose $\mathbf{F}'_1(\mathbf{s}, t) = \mathbf{F}'_2(\mathbf{s}, t)$, $\mathbf{M}(\mathbf{s}, t) = \beta_t$, and $\theta(\mathbf{s}, t) = \beta(\mathbf{s}, t)$ in our model. We then obtain the same mean function as Gelfend et al.. However, the main difference between their model and ours lays in the short–

term or small–scale spatial–temporal term. In their approach, they assume small–scale spatial–temporal term to be composed by two parts: a coregionalization spatial–temporal term and a measurement error component. However, Gelfand et al.'s model cannot deal with autocorrelated detrended residuals, and so is not applicable in our applications. Our model contains more general structure than theirs, and we actually decompose the short–term spatial–temporal term into two parts: the principal spatial pattern and a local spatial–temporal pattern.

7.3.4 Relationship with the BSP model in Le and Zidek (1992)

Our model also contains the Le–Zidek model as a special case. Le and Zidek prefilter the linear and seasonal trends, as well as the highly autocorrelated detrended residuals before fitting their model. This prefiltering step can be incorporated through the term $\mathbf{F}'_1(\mathbf{s}, t)\mathbf{M}(\mathbf{s}, t)$ and part of $\mathbf{F}'_2(\mathbf{s}, t)$ in our model. Removing (7.2), (7.4) and (7.5), we obtain the Le–Zidek BSP model.

Following (7.2), we can model the small spatial–temporal variation term as Le and Zidek do in their book. Let $W(\mathbf{s}, t) = V(\mathbf{s}, t) - \tilde{Z}(\mathbf{s}, t)\beta(\mathbf{s}, t)$ and $X(\mathbf{s}, t) = W(\mathbf{s}, t) - \phi_i W(\mathbf{s}, t - 1)$. For simplicity, assume $\beta(\mathbf{s}, t) = \beta_t$ and $\tilde{Z}(\mathbf{s}, t)$, the l_2 dimensional site–specific covariates vector, for example hourly temperatures or hourly wind speeds when the response variate is the hourly ozone concentrations in our study. We can write $\mathbf{W}_t = (W(\mathbf{s}_1, t), \ldots, W(\mathbf{s}_p, t))'$, and $\tilde{\mathbf{Z}}_t = (\tilde{Z}(\mathbf{s}_1, t)', \ldots, \tilde{Z}(\mathbf{s}_p, t)')' : p \times l_2$. We then have $\mathbf{W}_t = \mathbf{V}_t - \tilde{\mathbf{Z}}_\mathbf{t}\beta_t$ and $\mathbf{X}_t = \mathbf{W}_t - \mathrm{diag}(\phi_1, \ldots, \phi_p)\mathbf{W}_{t-1}$. Le and Zidek then model $\mathbf{X}'_t$ in a hierarchical Bayesian framework such that $\mathbf{X}'_t \sim N_p(\hat{\mathbf{z}}_t\mathbf{B}, \boldsymbol{\Sigma}^\mathrm{F})$, where $\hat{\mathbf{z}}_t$ represents a q–dimension covariate vector and $\mathbf{B} : q \times p$, the coefficient matrix. At the second level, Le and Zidek model $\mathbf{B} \sim N_{p \times q}(\mathbf{B}_0, \mathbf{F}^{-1} \otimes \boldsymbol{\Sigma}^\mathrm{F})$. Furthermore, they assume that $\boldsymbol{\Sigma}^\mathrm{F} \sim W_p^{-1}(\boldsymbol{\Psi}, \delta)$.

In our model, we also assume that $\beta_t = \mathbf{H}^\beta \beta_{t-1} + \omega_t^\beta$, where $\omega_t^\beta \sim N(\mathbf{0}, \mathbf{W}_2)$. We now have that

$$V(\mathbf{s}, t) = (\tilde{Z}(\mathbf{s}, t)\mathbf{H}^\beta - \phi_i \tilde{Z}(\mathbf{s}, t-1), \phi_i)(\beta'_{t-1}, V(\mathbf{s}, t-1))' + \tilde{Z}(\mathbf{s}, t)\omega_t^\beta$$

$$+X(\mathbf{s},t)$$
$$= \mathbf{F}_2'(\mathbf{s},t)\theta(\mathbf{s},t) + \tilde{Z}(\mathbf{s},t)\omega_t^\beta + X(\mathbf{s},t), \tag{7.27}$$

where $\mathbf{F}_2'(\mathbf{s},t) = (\tilde{Z}(\mathbf{s},t)\mathbf{H}^\beta - \phi_i\tilde{Z}(\mathbf{s},t-1), \phi_i) : 1 \times (l_2 + 1)$ and $\theta(\mathbf{s},t) = (\beta_{t-1}', V(\mathbf{s},t-1))' : (1 + l_2 \times 1)$.

By the Le–Zidek's modelling approach, we then have $\mathbf{V}_t = \mathbf{F}_{2t}'\theta_t + \tilde{\mathbf{Z}}_t\omega_t^\beta + \mathbf{B}\hat{\mathbf{z}}_t + \tau_t$, where $\tau_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}^{\mathrm{F}})$. Hence we obtain $\nu_t = \tilde{\mathbf{Z}}_t\omega_t^\beta + \tau_t \sim N(\mathbf{0}, \tilde{\mathbf{Z}}_t\mathbf{W}_2\tilde{\mathbf{Z}}_t' + \boldsymbol{\Sigma}^{\mathrm{F}})$. Note that this is a special case of our model.

7.4 A Multivariate Bayesian Spatio–Temporal Model

We now extend our model to a multivariate case. Suppose m different pollutants or species are measured in a spatio–temporal field. Let $Z_j(\mathbf{s},t)$ be the jth pollutant at site s and time t, for $j = 1,\dots,m$, $\mathbf{s} \in \{\mathbf{s}_1,\dots,\mathbf{s}_p\}$, and $t = 1,\dots,n$. Let $\mathbf{Z}(\mathbf{s,t}) = (Z_1(\mathbf{s},t),\dots,Z_m(\mathbf{s},t))'$ be the vector of observations at site $\mathbf{s}$ and time t. Assume l_{1j} type I covariates and l_{2j} type II covariates considered, for $j = 1,\dots,m$. Those covariates are represented by $\mathbf{F}_{1,j}(\mathbf{s},t)' : 1 \times l_{1j}$ and $\mathbf{F}_{2,j}(\mathbf{s},t)' : 1 \times l_{2j}$, respectively. Write $\mathbf{F}_i(\mathbf{s},t)' : m \times \sum_{j=1}^m l_{ij}$ as a block diagonal matrix with diagonal entries $\{\mathbf{F}_{i,1}(\mathbf{s},t)',\dots,\mathbf{F}_{i,m}(\mathbf{s},t)'\}$, for $i = 1,2$. Let $\mathbf{M}(\mathbf{s},t) = (\mathbf{M}_1(\mathbf{s},t)',\dots,\mathbf{M}_m(\mathbf{s},t)')' : \sum_{j=1}^m l_{1j} \times 1$ and $\theta(\mathbf{s},t) = (\theta_1(\mathbf{s},t)',\dots,\theta_m(\mathbf{s},t)')' : \sum_{j=1}^m l_{2j} \times 1$. Assume we obtain the $\boldsymbol{\Phi}_j^{K_j} : 1 \times K_j$, for $j = 1,\dots,m$, and the corresponding expansion coefficients $\mathbf{a}_j^{K_j}(t) : K_j \times 1$, using the multivariate EOF analysis. Let $\boldsymbol{\Phi}^*(\mathbf{s}) = \mathrm{Block\text{–}diag}\{\boldsymbol{\Phi}_1^{K_1}(\mathbf{s}),\dots,\boldsymbol{\Phi}_m^{K_m}(\mathbf{s})\} : m \times \sum_{j=1}^m K_j$, and $\mathbf{a}^*(t) = (\mathbf{a}_1^{K_1}(t),\dots,\mathbf{a}_m^{K_m}(t))' : \sum_{j=1}^m K_j \times 1$. We obtain the multivariate Bayesian spatio–temporal model as follows:

$$\mathbf{Z}(\mathbf{s},t) = \mathbf{F}_1'(\mathbf{s},t)\mathbf{M}(\mathbf{s},t) + \mathbf{Y}(\mathbf{s},t) \tag{7.28}$$
$$\mathbf{Y}(\mathbf{s},t) = \boldsymbol{\Phi}^*(\mathbf{s})\mathbf{a}^*(t) + \mathbf{V}(\mathbf{s},t) \tag{7.29}$$
$$\mathbf{V}(\mathbf{s},t) = \mathbf{F}_2'(\mathbf{s},t)\theta(\mathbf{s},t) + \nu(\mathbf{s},t) \tag{7.30}$$

and

$$\mathbf{M}(\mathbf{s},t) = \mathbf{G}_1(\mathbf{s},t)\mathbf{M}(\mathbf{s},t-1) + \eta(\mathbf{s},t) \qquad (7.31)$$

$$\mathbf{a}^*(t) = H_{2t}\mathbf{a}^*(t-1) + \xi_t \qquad (7.32)$$

$$\theta(\mathbf{s},t) = \mathbf{G}_2(\mathbf{s},t)\theta(\mathbf{s},t-1) + \omega(\mathbf{s},t), \qquad (7.33)$$

where $\mathbf{Y}(\mathbf{s},t) = (Y_1(\mathbf{s},t),\ldots,Y_m(\mathbf{s},t))' : m \times 1$, $\mathbf{V}(\mathbf{s},t) = (V_1(\mathbf{s},t),\ldots,$
$V_m(\mathbf{s},t))' : m \times 1$, $\nu(\mathbf{s},t) = (\nu_1(\mathbf{s},t),\ldots,\nu_m(\mathbf{s},t))' : m \times 1$, and $\omega(\mathbf{s},t) = (\omega_1(\mathbf{s},t),\ldots,\omega_m(\mathbf{s},t))' : m \times 1$. We assume a separable space–time covariance
structure for the matrix–variate $\nu_t = (\nu(\mathbf{s}_1,t),\ldots,\nu(\mathbf{s}_p,t)) : m \times p$, that is,

$$\nu_t \sim N_{m \times p}(\mathbf{0}, \mathbf{\Omega} \otimes \mathbf{\Sigma}^{\mathrm{F}}),$$

where $\mathbf{\Omega}$ represents the correlation matrix between the pollutants and $\mathbf{\Sigma}^{\mathrm{F}}$,
the covariance matrix between the spatial locations at the local– or fine–
scale level. We assume that the EOFs in $\mathbf{\Phi}^*$ estimate the covariance matrix,
that is, $\mathbf{\Omega} \otimes \mathbf{\Sigma}^{\mathrm{C}}$.

We also assume inverted Wishart priors for $\mathbf{\Sigma}^{\mathrm{F}}$ and $\mathbf{\Sigma}^{\mathrm{C}}$, respectively. We
leave that extension to future work.

7.5 MCMC Algorithm on the Bayesian Spatio–temporal Models

In this section, we discuss MCMC algorithms for sampling from the joint
posterior distributions of the model parameters for both univariate and mul-
tivariate spatial–temporal models. We do not present a simulation study or
application, pending completion of future work on implementation. Instead
we stop with a fairly clear statement of an approach that makes success
seem plausible in applications like that one presented earlier to the AQS
data, with hourly ozone concentrations over the entire USA and the whole
of one summer.

To make computation feasible, we can estimate the K using an ad hoc
method. Or we can implement the idea of Bayesian EOF Chapter 6 in this

new proposed model, using block MCMC methods to draw samples from the full conditional distributions of model parameters.

As we can see from the prediction posterior distribution for the univariate Bayesian spatio–temporal model, the Metropolis–Hasting method is required to sample from the posterior distributions of the the fine– and coarse–level covariance matrices. That allows the Generalized inverted Wishart (GIW) prior to work at this stage. We leave this extension in future work.

We now summarize what we have obtained in this chapter. From the predictive posterior distribution obtained in Section 7.2, we can use the block MCMC method to draw samples from its joint posterior distribution. In other words, we iteratively sample from $p(\boldsymbol{\Sigma}^{\mathrm{F}}|\boldsymbol{\Sigma}^{\mathrm{C}}, \mathbf{Z}_{1:n})$, $p(\boldsymbol{\Sigma}^{\mathrm{C}}|\boldsymbol{\Sigma}^{\mathrm{F}}, \mathbf{Z}_{1:n})$, and $p(\theta_{1:n}^{*}|\boldsymbol{\Sigma}^{\mathrm{F}}, \boldsymbol{\Sigma}^{\mathrm{C}}, \mathbf{Z}_{1:n})$ according to (7.25), (7.26), and Theorem 7.2.1, respectively.

Algorithm 7.5.1 *(Metropolis–within–Gibbs algorithm)*

1. *Initialization: sample*

$$
\begin{aligned}
\boldsymbol{\Sigma}^{-F(1)} &\sim W(\boldsymbol{\Xi}^{F}, \delta^{F}) \\
\boldsymbol{\Sigma}^{-C(1)} &\sim W(\boldsymbol{\Xi}^{C}, \delta^{C}) \\
\theta_{1:n}^{*}{}^{(1)} &\sim N(\boldsymbol{m}_{0}, \boldsymbol{C}_{0}).
\end{aligned}
$$

2. *Given the $(j-1)^{th}$ values, $\boldsymbol{\Sigma}^{-F(j-1)}$, $\boldsymbol{\Sigma}^{-C(j-1)}$, $\theta_{1:n}^{*}{}^{(j-1)}$ and the observations $\boldsymbol{Z}_{1:n}$:*

    (1) Use a Metropolis–Hasting step to sample $\boldsymbol{\Sigma}^{-F(j)}$ from $p(\boldsymbol{\Sigma}^{-F}|\boldsymbol{\Sigma}^{-C}, \theta_{1:n}^{*}, \boldsymbol{Z}_{1:n})$.

    (2) Use a Metropolis–Hasting step to sample $\boldsymbol{\Sigma}^{-C(j)}$ from $p(\boldsymbol{\Sigma}^{-C}|\boldsymbol{\Sigma}^{-F}, \theta_{1:n}^{*}, \boldsymbol{Z}_{1:n})$.

    (3) Sample $\theta_{1:n}^{*}{}^{(j)}$ from $p(\theta_{1:n}^{*}|\boldsymbol{\Sigma}^{-C}, \boldsymbol{\Sigma}^{-F}, \boldsymbol{Z}_{1:n})$.

3. *Repeat until convergence.*

The implementation of the MCMC method to sample the model parameters in our model will be left to future work. After obtaining the MC samples

for the model parameters after the burn–in period, it is straightforward to obtain the temporal prediction of the responses using the DLM approach. The spatial interpolation problem can then be viewed as a missing data in the DLM. The latter can be obtained by adding this MCMC block of missing data in the above algorithm. We leave all the implementation of this model to future wok.

7.6   Results and Conclusions

Our Bayesian spatio–temporal model is very flexible and powerful. It has the following advantages:

1. The general structure of the unified Bayesian spatio–temporal model allows us to remove the long–term systematic temporal variation, for example, the linear trend, seasonal trends, covariate effects for both types. For type I covariates, that is, the site–specific covariates, we regress them on site–invariant but time–varying unknown coefficients, an extension of the Le–Zidek's approach. For type II covariates, that is, time–varying covariates common over the entire domain, we regress them on a site–specific and time–varying unknown coefficient matrix, one removed at the process model in Le–Zidek's approach.

2. We then decompose $Y(\mathbf{s}, t)$ into two parts: the principal local spatial patterns and the remaining local spatial–temporal variation terms. In this step, $\mathbf{\Phi}^K(\mathbf{s})$ can also be constructed using other basis functions, such as orthogonal polynomials, thin spline plates, and bicubic spline methods. However, the eigenvalue basis functions, that is, the basis functions for the EOF, do provide a unique solution by the spectral decomposition theorem. In other words, we can truncate the first K EOFs to represent the main spatial patterns in the spatial–temporal fields of $Y(\mathbf{s}, t)$, that is, the "detrended" residuals.

3. The remaining local spatial–temporal variation term is then modelled as a BSP term because of the need for computational feasibility. The

general structure of the systematic temporal components can then be incorporated. From this point of view, we obtain an extension for the SSM proposed by Wikle and Cressie (1999).

4. Our model allows us to incorporate two different scales of covariance matrices, that is, coarse– and fine– scale levels of spatial covariance matrices. At the coarse level, we truncate the EOFs using the Bayesian EOF method in Chapter 6 to represent the coarse scale long–term spatial patterns. At the fine level, the BSP term is used to represent the local scale short–term spatial–temporal patterns.

5. Spatial interpolation and temporal prediction turn out to be straight-forward using our model. Gaussian framework allows us to spatially interpolate the responses at ungauged sites. The time evolution of the dynamic models allows us to temporally predict the responses at the gauged and ungauged sites.

Future work includes implementing that model to the real database or simulation study, and an extension involving the GIW prior for the coarse– and fine– scale covariance matrices.

Chapter 8

Future Work

This last chapter presents a summary of this thesis (Section 8.1) and proposals for future work based on it (Section 8.2) as well as approaches to be taken to complete that work.

8.1   Thesis Summary

We have implemented the Gaussian DLM proposed by Huerta et al. (2004) to spatially interpolate ground–level ozone concentrations at one cluster of monitoring stations in an AQS database (1995). A complete and tested software package has been developed for this implementation. Theoretical results have been developed regarding the predictive variance using the first–order polynomial model. Those results explain the monotone behavior of the coverage probabilities found from the results of spatial interpolation. Moreover, we demonstrate how to use discount factors in the DLM to improve the predictive results and their accuracy. However, we find the approach very computationally intensive so that it is not scalable to large space–time domains.

We therefore explore a computational simpler alternative, the BSP approach, for spatially interpolating univariate and multivariate responses in space–time domains. That alternative has been implemented for an AQS database (2000). Moreover, we have found an extension to this approach for temporal forecasting one–day–ahead ground–level ozone concentrations. After comparing it with the DLM, we find that the latter cannot compete with the BSP for either spatial interpolation or temporal prediction in terms of mean squared prediction error (MSPE).

The BSP approach uses empirical Bayes steps to estimate some model

parameters within the Bayesian framework. Although these steps simplify computation the *ad hoc* approach would be seen as objectionable from a purely Bayesian perspective and the model is not as flexible as the DLM. To refine the BSP therefore, we put it into a DLM framework so that the model is more flexible to update or predict the responses as new data come on stream, one of the attractive features of dynamic models. Yet we preserve its computational simplicity.

A Bayesian version of the EOF method, that is, the BEOF method, has been proposed in this thesis. The BEOF method allows us to represent the principal spatial patterns of spatial–temporal fields in a fully Bayesian framework. We have demonstrated potentially severe problems with the classical EOF method that seems to have been largely ignored. We have compared the classical and corrected EOFs with true EOFs using the simulated database. Moreover, we have developed an MCMC algorithm for sampling from the joint posterior distribution of model parameters. However, implementation of the BEOF approach on our existing database will be left to future work.

Finally, we have proposed a unified approach to univariate and multivariate spatio–temporal modelling within a fully hierarchical Bayesian framework so that we can incorporate some interesting features of the BSP, DLM and BEOF approaches. We have provided theoretical results on the joint posterior distribution of model parameters and the corresponding MCMC algorithm. Implementation of this model in data and simulated data will be completed in future work.

8.2   Future Research Plan

We finally propose in the list below, future work based on this thesis as well as possible directions for completing that work.

1. We can extend the DLM to the discount DLM and implement it on a real database. This implementation helps us set the reasonable hyperpriors for some of the model parameters instead of fixing them in

an ad–hoc way. We can also develop corresponding software for this purpose.

2. We can extend the DLM to incorporate the dependence of model parameters. To do this, a further extension of the discount DLM is needed so that the time–varying discount factors can be free of the computational burden.

3. We will extend the one–day–ahead temporal prediction of ground–level ozone concentrations using the BSP approach to arbitrary time points.

4. One interesting problem in the BSP approach is to deal with the monotone data (double staircase) patterns in two directions. This is a really difficult problem and a route to the solution remains to be found.

5. We can implement the BSP in multivariate cases where dependent but unknown structure between the multivariate responses are considered.

6. Another future project for the BSP approach involves misaligned data. This problem will be partially addressed when we integrate the BSP into the DLM framework, so the model we considered in Chapter 7 will be one possible choice. The DLM framework helps in the prediction of missing data provided the rate of missingness is reasonable.

7. We will explore the choice of an optimal starting hour for making one–day–ahead prediction.

8. The Bayesian EOF method needs to be applied to real as opposed to simulated data so that we can find whether the Bayesian and corrected EOFs can give seriously discrepant answers, a result that would raise concerns about the appropriateness of classical EOFs. This implementation is straightforward following the MCMC algorithm we developed.

9. The extended BSP, i.e., the Bayesian spatio–temporal model, can be implemented with either model output or measurements. By combining the two types of data over the entire network of US EPA moni-

toring sites, more accurate spatial predictions should be achievable for ground–level ozone concentrations over a large spatio–temporal domain. This straightforward implementation of our theory follows from our theoretical results and MCMC algorithm we have already developed. For example, we can combine physical model outputs with statistical ones, in which the MAQSIP (gridded) data, can be considered at the regional or coarse level and the observations or measurements at each monitoring station, at the local or fine level.

10. We can extend Bayesian spatio–temporal models to incorporate different prior beliefs for the model parameters, that is, the covariance matrices regarding to different levels and blocks. This extension includes developing the corresponding theoretical results on predictive and posterior distributions using the GIW prior on model parameters.

11. We can use MCMC sampling algorithm to sample a random covariance matrix that follows a GIW prior distribution. This algorithm will help us to spatially interpolate and temporally predict the responses using the Bayesian spatio–termporal models in which the priors of covariance matrices at both coarse and fine levels are GIWs.

Bibliography

[1] Björnsson, H. & Venegas, S.A. (1997). *A manual for EOF and SVD analyses of climatic data.* Centre for Climate and Global Change Research Report # 97–1, McGill U.

[2] Bretthorst, L. (1988). *Bayesian spectral analysis and estimation.* New York: Springer Verlag.

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

[4] Brown, P.J., Le, N.D., & Zidek, J.V. (1994b). Inference for a covariance matrix. In *Aspects of uncertainty: A tribute to DV Lindly, (Eds) PR Freeman and AFM Smith.* Chichester: Wiley, 77–92.

[5] Calder, C.A. (2004). Efficient posterior inference and prediction of space–time processes using dynamic process convolutions. In the *Joint Proceedings of the Sixth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences and the Fifteenth Annual Conference of TIES.* Portland, Maine. June 28–July 1, 2004.

[6] Calder, C.A., & Cressie, N. (2007). Some topics in convolution–based spatial modeling. *Proceedings of the 56th Session of the International Statistics Institute.* Lisbon, Portugal. August 22–29, 2007.

[7] Carter, C.K., & Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika,* 81, 541–553.

[8] Chikuse Y. (2006). State space models on special manifolds. *Journal of Multivariate Analysis,* 97, 1284–1294.

[9] Cressie, N. (1993). *Statistics for spatial data.* Wiley.

[10] Cressie N., & Wikle, C.K. (1998). Strategies for dynamic space–time statistical modelling: discussion of "The Kriged Kalman filter" by Mardia et. al. *Test,* 7, 257–264.

[11] Damian, D., Sampson, P.D., & Guttorp, P. (2001). Bayesian estimation of semi–parametric non–stationary spatial covariance structures. *Environmetrics,* 12, 161–178.

[12] Dou, Y.P., Nhu, D.L., & Zidek J.V. (2007). *A dynamic linear model for hourly ozone concentrations.* Technical Report #228, Department of Statistics, UBC.

[13] Fu, A., Le, N.D., & Zidek, J.V. (2003). A statistical characterization of a simulated Canadian annual maximum rainfall field. Technical report #2003–17, SAMSI.

[14] Fuentes, M. (2002). Interpolation of nonstationary air pollution process: a spatial spectral apporach. *Statistical Modelling,* 2, 281–298.

[15] Gamerman, D., & Lopes, H.F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian infrence.* Chapman & Hall.

[16] Gelfand, A.E., Banerjee, S., & Gamerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics,* 16, 465–479.

[17] Gelfand, A.E., Schmidt, A.M., Banerjee, S., & Sirmans, C.F. (2004). Nonstationary multivariate process modelling through spatially varying coregionalization (with discussion). *Test,* 13, 263–312.

[18] Handcock, M.S., & Stein M.L. (1993). A Bayesian analysis of kriging. *Technometrics,* 35, 403–410.

[19] Hannachi, A., Jolliffe, I.T., & Stephenson, D.B. (2007). Review empirical orthogonal functions and related techniques in atmospheric science: a review *Int. J. Climatol.,* 27, 1119–1152.

[20] Harvey, A.C. (1984). A unified view of statistical forecasting procedures. *J. Forecasting,* 3, 245–275.

[21] Higdon, D., Swall, J., & Kern, J. (1998). Non–stationary spatial modeling. *Bayesian statistics 6,* eds. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith. Oxford: Oxford University Press, 761–768.

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

[23] James A.T. (1954). Normal multivariate analysis and the orthogonal group.

[24] James, A.T. (1960). The distribution of the latent roots of the covariance matrix. *The Annals of Mathematical Statistics,* 31, 151–158.

[25] Johannesson, G., Cressie, N., & Huang, H. (2007). Dynamic multi–resolution spatial models. *Environmental and Ecological Statistics,* 14, 5–25.

[26] Kyriakidis, P.C., & Journel, A.G. (1999). Geostatistical space–time models: A review. *Mathematical Geology,* 31, 651–684.

[27] Le, N.D., Sun, L., & Zidek, J.V. (2001). Spatial prediction and temporal backcasting for environmental fields having monotone data patterns. *Can. J. Statist.,* 29, 516–529.

[28] Le, N.D., Sun, W., & Zidek, J.V. (1997). Bayesian multivariate spatial interpolation with data missing by design. *J. Roy. Stat. Soc.,* Ser B, 59, 501–510.

[29] Le, N.D., & Zidek, J.V. (1992). Interpolation with uncertain spatial covariance: A Bayesian alternative to kriging. *J. Mult. Anal.,* 43, 351–374.

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

[31] Lee, H., & Ghosh, S.K. (2005). A reparametrization approach for dynamic space–time models. *Institude of Statistics Mimeo Series # 2587.*

[32] Li, K.H., Le, N.D., Sun, L., & Zidek, J.V. (1999). Spatial–temporal models for ambient hourly PM_{10} in Vancouver. *Environmetrics,* 10, 321–338.

[33] Lopes, H.F., Salazar, E., & Gamerman, D. (2006). Spatial dynamic factor analysis. *Techical report, Universidade Federal do Rio de Janeiro.*

[34] Mardia, K.V. (1977). Distributions on Stiefel and Grassman manifolds, and their applications. *Advances in Applied Probability,* 9, 435–436.

[35] Mardia, K.V., Goodall, C., Redfern, E.J., & Alonso, F.J. (1998). The kriged Kalman filter. *Test,* 7, 217–282.

[36] Mardia, K.V., & Khatri, C.G. (1977). Uniform distribution on a Stiefel manifold. *Journal of Multivariate Analysis,* 7, 468–473.

[37] Moller, J. (2003). *Spatial statistics and computational methods.* Springer–Verlag.

[38] Sampson, P., & Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial structure. *J. Amer. Stat. Assoc.,* 87, 108–119.

[39] Stroud, J.R., Muller, P., & Sanso, B. (2001). Dynamic models for spatio–temporal data. *J. R. Statist. Soc. B,* 63, 673–689.

[40] Sun, W., Le, N.D., Zidek, J.V., & Burnett, R. (1998). Assessment of Bayesian multivariate interpolation approach for health impact studies. *Environmetrics,* 9, 565–586.

[41] Wackernagel, H. (1998) *Multivariate geostatistics: an introduction with applications.* Springer.

[42] West, M., & Harrison, J. (1997). *Bayesian forcasting and dynamic models.* Springer–Verlag.

[43] Wikle, C.K. (2002). Spatio–temporal methods in climatology. *To appear: Encyclopedia of Life Support Systems.* EOLSS Publishers Co. Ltd.

[44] Wikle, C.K., & Cressie, N. (1999). A dimension–reduced approach to space–time Kalman filtering. *Biometrika,* 86, 815–829.

[45] Zidek, J.V., Sun, L., Le, N.D., 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.

Appendix A

Additional Results for Chapter 2

A.1 Additional Results for Section 2.6.1

The joint posterior distribution for $\mathbf{x}_{1:T}, \lambda$ and σ^2 is given by

$$
\begin{aligned}
p(\mathbf{x}_{1:T}, \lambda, \sigma^2 | \mathbf{y}_{1:T}) &= p(\lambda, \sigma^2) p(\mathbf{x}_T | \lambda, \sigma^2, \mathbf{y}_{1:T}) \prod_{t=1}^{T} p(\mathbf{x}_{T-t} | \mathbf{x}_{T-t+1}, \lambda, \sigma^2, \mathbf{y}_{1:T}) \\
&\quad \times \prod_{t=1}^{T} p(\mathbf{y}_t | \lambda, \sigma^2, \mathbf{y}_{1:t-1}) \\
&= p(\mathbf{x}_{1:T} | \lambda, \sigma^2, \mathbf{y}_{1:T}) p(\sigma^2 | \lambda, \mathbf{y}_{1:T}) p(\lambda | \mathbf{y}_T).
\end{aligned}
$$

Suppose $p(\lambda, \sigma^2) = p(\lambda)p(\sigma^2)$, that is, the priors for λ and σ^2 are independent with each other.

The joint posterior distribution for λ and σ^2 can be written as follows:

$$
p(\lambda, \sigma^2 | \mathbf{y}_{1:T}) \propto p(\lambda) p(\sigma^2) (\sigma^2)^{-nT/2} \prod_{t=1}^{T} |\mathbf{Q}_t|^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^{T} \mathbf{e}_t' \mathbf{Q}_t^{-1} \mathbf{e}_t \right\}.
$$

If the prior for σ^2 is an inverse gamma distribution with shape parameter α and scale parameter β, then the posterior distribution for σ^2 is also an inverse gamma distribution with shape parameter $\alpha + \frac{nT}{2}$ and scale parameter $\beta + \frac{1}{2} \sum_{t=1}^{T} \mathbf{e}_t' \mathbf{Q}_t^{-1} \mathbf{e}_t$.

Hence, the posterior density for λ can be written as follows:

$$
p(\lambda | \mathbf{y}_{1:T}) = \frac{p(\lambda, \sigma^2 | \mathbf{y}_{1:T})}{p(\sigma^2 | \lambda, \mathbf{y}_{1:T})}
$$

$$\propto \quad p(\lambda)\prod_{t=1}^{T}|\mathbf{Q}_t|^{-1/2}\left[\beta+\frac{1}{2}\sum_{t=1}^{T}\mathbf{e}_t'\mathbf{Q}_t^{-1}\mathbf{e}_t\right]^{-(\alpha+nT/2)}.$$

Therefore, the posterior density for $\mathbf{x}_{1:T}$ is given by

$$p(\mathbf{x}_{1:T}|\lambda,\sigma^2,\mathbf{y}_{1:T}) \quad = \quad p(\mathbf{x}_T|\lambda,\sigma^2,\mathbf{y}_{1:T})\prod_{t=1}^{T}p(\mathbf{x}_{T-t}|\mathbf{x}_{T-t+1},\lambda,\sigma^2,\mathbf{y}_{1:T}).$$

A.2   Additional Results for Section 2.6.2

Theorem A.2.1 *Under Models (2.14) - (2.15) and initial prior (2.18), for any $1 \le t \le T$, conditional on θ, we have*

(i)

$$
\begin{aligned}
(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1},\theta) &\sim N[\boldsymbol{m}_{t-1},\sigma^2\boldsymbol{C}_{t-1}] \\
(\boldsymbol{x}_{t}|\boldsymbol{y}_{1:t-1},\theta) &\sim N[\boldsymbol{a}_t,\sigma^2\boldsymbol{R}_t] \\
(\boldsymbol{y}_{t}|\boldsymbol{y}_{1:t-1},\theta) &\sim N[\boldsymbol{f}_t,\sigma^2\boldsymbol{Q}_t] \\
(\boldsymbol{x}_{t}|\boldsymbol{y}_{1:t},\theta) &\sim N[\boldsymbol{m}_t,\sigma^2\boldsymbol{C}_t],
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{a}_t &= \boldsymbol{m}_{t-1} & \boldsymbol{R}_t &= \boldsymbol{C}_{t-1}+\boldsymbol{W} \\
\boldsymbol{f}_t &= \boldsymbol{F}_t'\boldsymbol{a}_t & \boldsymbol{Q}_t &= \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t+\boldsymbol{V}_\lambda \\
\boldsymbol{e}_t &= \boldsymbol{y}_t-\boldsymbol{f}_t & \boldsymbol{A}_t &= \boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{Q}_t^{-1} \\
\boldsymbol{m}_t &= \boldsymbol{a}_t+\boldsymbol{A}_t\boldsymbol{e}_t & \boldsymbol{C}_t &= \boldsymbol{R}_t-\boldsymbol{A}_t\boldsymbol{Q}_t\boldsymbol{A}_t'.
\end{aligned}
$$

(ii) Define $\boldsymbol{B}_t = \boldsymbol{C}_t\boldsymbol{R}_{t+1}^{-1}$. For $0 \le k \le T-1$,

$$(\boldsymbol{x}_{T-k}|\boldsymbol{y}_{1:T},\theta) \quad \sim \quad N[\boldsymbol{a}_T(-k),\sigma^2\boldsymbol{R}_T(-k)], \tag{A.1}$$

where

$$
\begin{aligned}
\boldsymbol{a}_T(-k) &= \boldsymbol{m}_{T-k}+\boldsymbol{B}_{T-k}[\boldsymbol{a}_T(-k+1)-\boldsymbol{a}_{T-k+1}] \\
\boldsymbol{R}_T(-k) &= \boldsymbol{C}_{T-k}+\boldsymbol{B}_{T-k}[\boldsymbol{R}_T(-k+1)-\boldsymbol{R}_{T-k+1}]\boldsymbol{B}_{T-k}',
\end{aligned}
$$

with $\boldsymbol{a}_T(0) = \boldsymbol{m}_T$, $\boldsymbol{R}_T(0) = \boldsymbol{C}_T$, $\boldsymbol{a}_{T-k}(1) = \boldsymbol{a}_{T-k+1}$, and $\boldsymbol{R}_{T-k}(1) = \boldsymbol{R}_{T-k+1}$.

A.3 Additional Results for Section 2.6.4

The observation equation is given by

$$\mathbf{y}_t \;\; = \;\; \mathbf{1}_n'\beta_t + S_{1t}(a_1)\alpha_{1t} + S_{2t}(a_2)\alpha_{2t} + \nu_t, \quad \nu_t \sim N[\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}(\lambda)],$$

where $\boldsymbol{\Sigma}(\lambda) = \exp(-\mathbf{V}/\lambda)$ and $\mathbf{V}$ denotes the distance matrix for the monitoring sites $\mathbf{s}_1, \ldots, \mathbf{s}_n$.

Given λ, σ^2, $\mathbf{x}_t$ (that is, β_t, α_{1t} and α_{2t}) and $\mathbf{y}_t$, the posterior conditional distribution for the constant phase parameters, a_1 and a_2, is given by

$$
\begin{aligned}
p(a_1, a_2 | \lambda, \sigma^2, \mathbf{x}_t, \mathbf{y}_t) \;\; &\propto \;\; p(\mathbf{y}_t | \lambda, \sigma^2, \mathbf{x}_t, a_1, a_2) p(a_1, a_2) \\
&\propto \;\; p(\mathbf{M}_t | \lambda, \sigma^2, \beta_t, \alpha_{1t}, \alpha_{2t}, a_1, a_2) p(a_1, a_2),
\end{aligned}
$$

where $\mathbf{M}_t = \mathbf{y}_t - \mathbf{1}_n'\beta_t - \cos(\frac{\pi t}{12})\alpha_{1t} - \cos(\frac{\pi t}{6})\alpha_{2t}$, for $t = 1, \ldots, T$.

We consider the following two cases for the prior of $\mathbf{a} = (a_1, a_2)'$:

- *Case (i)* a standard reference prior: $p(\mathbf{a}) \propto 1$;

- *Case (ii)* a bivariate normal prior: $\mathbf{a} \sim N(\mu, \boldsymbol{\Sigma})$, where $\mu = (\mu_1, \mu_2)'$ and $\boldsymbol{\Sigma}$ is a 2 by 2 covariance matrix.

Under Case (i), for fixed $t = 1, \ldots, T$, let $\mathbf{l}_1 = \sin(\frac{\pi t}{12})\alpha_{1t}$, $\mathbf{l}_2 = \sin(\frac{\pi t}{6})\alpha_{2t}$, $\mathbf{m} = \mathbf{M}_t$ and $\mathbf{S} = \sigma^2 \boldsymbol{\Sigma}(\lambda)$. The posterior conditional distribution for the phase parameter vector $\mathbf{a}$ is now given by

$$
\begin{aligned}
p(a_1, a_2 | \lambda, \sigma^2, \mathbf{x}_t, \mathbf{y}_t) \;\; &\propto \;\; p(\mathbf{M}_t | \lambda, \sigma^2, \mathbf{x}_t, a_1, a_2, \mathbf{x}_t, \lambda, \sigma^2) p(a_1, a_2) \\
&\propto \;\; \exp\{-1/2[\mathbf{M}_t - a_1 \sin(\frac{\pi t}{12})\alpha_{1t} - a_2 \sin(\frac{\pi t}{6})\alpha_{2t}]' \\
&\qquad \times (\sigma^2 \boldsymbol{\Sigma}(\lambda))^{-1}[\mathbf{M}_t - a_1 \sin(\frac{\pi t}{12})\alpha_{1t} - a_2 \sin(\frac{\pi t}{6})\alpha_{2t}]\} \\
&= \;\; \exp\{-\frac{1}{2}(\mathbf{m} - a_1 \mathbf{l}_1 - a_2 \mathbf{l}_2)'\mathbf{S}^{-1}(\mathbf{m} - a_1 \mathbf{l}_1 - a_2 \mathbf{l}_2)\}
\end{aligned}
$$

$$\propto \ \exp\{-\frac{1}{2}(\mathbf{a}\mathbf{\Sigma}^{-1}\mathbf{a}' - 2\mathbf{a}\mathbf{\Sigma}^{-1}\mu')\}$$

$$\propto \ \exp\{-\frac{1}{2}[\mathbf{a}'(\mathbf{l}_1,\mathbf{l}_2)'\mathbf{S}^{-1}(\mathbf{l}_1,\mathbf{l}_2)\mathbf{a} - \mathbf{a}'(\mathbf{l}_1,\mathbf{l}_2)'\mathbf{S}^{-1}\mathbf{m}$$
$$-\mathbf{m}S^{-1}(\mathbf{l_1},\mathbf{l_2})\mathbf{a}]\}$$

$$\propto \ \exp\{-\frac{1}{2}(\mathbf{a} - \mu)'\mathbf{\Sigma}^{-1}(\mathbf{a} - \mu)\},$$

where

$$\mathbf{\Sigma}^{-1} \ = \ (\mathbf{l}_1,\mathbf{l}_2)'\mathbf{S}^{-1}(\mathbf{l}_1,\mathbf{l}_2), \tag{A.2}$$

$$\mu \ = \ \mathbf{\Sigma}(\mathbf{l}_1,\mathbf{l}_2)'\mathbf{S}^{-1}\mathbf{m}. \tag{A.3}$$

Note that equation (A.3) is equivalent to

$$\mathbf{\Sigma}^{-1}\mu = (\mathbf{l}_1,\mathbf{l}_2)'\mathbf{S}^{-1}\mathbf{m}. \tag{A.4}$$

More specifically, we obtain the following elements in the mean vector and covariance matrix for the posterior conditional distribution of the phase parameter vector $\mathbf{a}$.

$$\mathbf{\Sigma} \ = \ \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}, \tag{A.5}$$

$$\Delta^{-1} \ = \ (\sigma_{11}\sigma_{22} - \sigma_{12}^2)^{-1} \tag{A.6}$$

$$= \ (\mathbf{l}_1'\mathbf{S}^{-1}\mathbf{l}_1)(\mathbf{l}_2'\mathbf{S}^{-1}\mathbf{l}_2) - (\mathbf{l}_1'\mathbf{S}^{-1}\mathbf{l}_2)^2 \tag{A.7}$$

$$\sigma_{11} \ = \ \Delta(\mathbf{l}_2'\mathbf{S}^{-1}\mathbf{l}_2) \tag{A.8}$$

$$\sigma_{12} \ = \ -\Delta(\mathbf{l}_1'\mathbf{S}^{-1}\mathbf{l}_2) \tag{A.9}$$

$$\sigma_{22} \ = \ \Delta(\mathbf{l}_1'\mathbf{S}^{-1}\mathbf{l}_1) \tag{A.10}$$

$$\mu_1 \ = \ \sigma_{11}(\mathbf{l}_1'\mathbf{S}^{-1}\mathbf{m}) + \sigma_{12}(\mathbf{l}_2'\mathbf{S}^{-1}\mathbf{m}) \tag{A.11}$$

$$\mu_2 \ = \ \sigma_{12}(\mathbf{l}_1'\mathbf{S}^{-1}\mathbf{m}) + \sigma_{22}(\mathbf{l}_2'\mathbf{S}^{-1}\mathbf{m}) \tag{A.12}$$

Therefore, we have the following conclusions for the conditional posterior distribution of the phase parameter vector $\mathbf{a}$.

(i) If the prior for $\mathbf{a}$ is the standard reference prior, that is, $p(a_1, a_2) \propto 1$, we have

$$\left(\begin{array}{c} a_1 \\ a_2 \end{array} \bigg| \mathbf{x}_t, \mathbf{y}_t, \lambda, \sigma^2 \right) \sim N \left[\left(\begin{array}{c} \mu_1 \\ \mu_2 \end{array} \right), \mathbf{\Sigma} \right],$$

where μ_1, μ_2 and $\mathbf{\Sigma}$ can be formed in equations (A.5) - (A.12) or equations (A.2) - (A.3).

(ii) If the prior for $\mathbf{a}$ is the bivariate normal distribution with mean vector $\mu^{\mathbf{0}} = (\mu_1^0, \mu_2^0)'$ and covariance matrix

$$\mathbf{\Sigma}^0 = \left(\begin{array}{cc} \sigma_{11}^0 & \sigma_{12}^0 \\ \sigma_{12}^0 & \sigma_{22}^0 \end{array} \right),$$

we then have

$$
\begin{aligned}
p(a_1, a_2 | \lambda, \sigma^2, \mathbf{x}_t, \mathbf{y}_t) \quad \propto \quad & \exp\{-\frac{1}{2}(a_1 - \mu_1, a_2 - \mu_2)' \mathbf{\Sigma}^{-1}(a_1 - \mu_1, \\
& a_2 - \mu_2)\} \exp\{-\frac{1}{2}(a_1 - \mu_1^0, a_2 - \mu_2^0)'(\mathbf{\Sigma}^0)^{-1} \\
& \times (a_1 - \mu_1^0, a_2 - \mu_2^0)\} \\
\propto \quad & \exp\{-\frac{1}{2}(a_1 - \mu_1^*, a_2 - \mu_2^*)' \mathbf{\Sigma}^{*-1}(a_1 - \mu_1^*, \\
& a_2 - \mu_2^*)\},
\end{aligned}
$$

where

$$\mathbf{\Sigma}^* = (\mathbf{\Sigma}^{-1} + \mathbf{\Sigma}^{0-1})^{-1} \tag{A.13}$$

$$\mu^* = \mathbf{\Sigma}^*(\mathbf{\Sigma}^{-1}\mu + (\mathbf{\Sigma}^0)^{-1}\mu_{\mathbf{0}}). \tag{A.14}$$

From (A.13) and (A.14), we have

$$\mathbf{\Sigma}^* = \mathbf{\Sigma} - \mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{\Sigma}^0)^{-1}\mathbf{\Sigma} = \mathbf{\Sigma}^0(\mathbf{\Sigma} + \mathbf{\Sigma}^0)^{-1}\mathbf{\Sigma}, \tag{A.15}$$

and

$$\mu^* = \mathbf{\Sigma}^0(\mathbf{\Sigma} + \mathbf{\Sigma}^0)^{-1}\mu + \mathbf{\Sigma}(\mathbf{\Sigma} + \mathbf{\Sigma}^0)^{-1}\mu^{\mathbf{0}}. \tag{A.16}$$

Hence, the posterior conditional distribution for the phase parameters is given by

$$\left(\mathbf{a}|\lambda, \sigma^2, \mathbf{x}_t, \mathbf{y}_t\right) \sim N(\mu^*, \mathbf{\Sigma}^*),$$

where μ^* and $\mathbf{\Sigma}^*$ can be referred to equations (A.13)–(A.14), or (A.15)–(A.16).

A.4 Additional Results for Sections 2.7.1 and 2.7.2

Given the values of the phase parameters, range and variance parameters, and the observations until time t, the joint distribution of $\alpha_{1t}^s, \alpha_{1t}$ is

$$\left(\begin{array}{c} \alpha_t^s \\ \alpha_{1t} \end{array}\right) \sim \mathrm{N}\left[\left(\begin{array}{c} \alpha_{1,t-1}^s \\ \alpha_{1,t-1} \end{array}\right), \sigma^2 \tau_1^2 \mathbf{\Sigma}^*(\lambda_1),\right]$$

where

$$\mathbf{\Sigma}^*(\theta) = \exp\{-\mathbf{V}^*/\theta\} = \left[\begin{array}{cc} \Sigma_{11}^*(\theta) & \mathbf{\Sigma}_{12}^*(\theta) \\ \mathbf{\Sigma}_{21}^*(\theta) & \mathbf{\Sigma}_{22}^*(\theta) \end{array}\right],$$

with $\Sigma_{11}^*(\theta)$ a scalar, $\mathbf{\Sigma}_{12}^*(\theta)$ a 1 by n vector, and $\mathbf{\Sigma}_{22}^*(\theta)$ a n by n matrix. We use $\mathbf{V}^*$ to denote the new distance matrix for the unknown site $\mathbf{s}$ and the monitoring stations $\mathbf{s}_1, \ldots, \mathbf{s}_n$.

We then have the conditional posterior distribution of α_{1t}^s as follows:

$$\begin{aligned}
(\alpha_{1t}^s|\alpha_{1,t-1}^s, \alpha_{1t}, \alpha_{1,t-1}, \mathbf{y}_t, \lambda, \sigma^2) \quad \sim \quad & N[\alpha_{1,t-1}^s + \mathbf{\Sigma}_{12}^*(\lambda_1)\mathbf{\Sigma}_{22}^*(\lambda_1)^{-1}(\alpha_{1t} - \\
& \alpha_{1,t-1}), \sigma^2\tau_1^2(\Sigma_{11}^*(\lambda_1) - \mathbf{\Sigma}_{22}^*(\lambda_1)^{-1} \\
& \times\mathbf{\Sigma}_{21}^*(\lambda_1))].
\end{aligned}$$

$$\text{(A.17)}$$

Similarly, the conditional posterior distribution for α_{2t}^s is

$$\begin{aligned}
(\alpha_{2t}^s|\alpha_{2,t-1}^s, \alpha_{2t}, \alpha_{2,t-1}, \mathbf{y}_t, \lambda, \sigma^2) \quad \sim \quad & N[\alpha_{2,t-1}^s + \mathbf{\Sigma}_{12}^*(\lambda_2)\mathbf{\Sigma}_{22}^*(\lambda_2)^{-1}(\alpha_{2t} - \\
& \alpha_{2,t-1}), \sigma^2\tau_2^2(\Sigma_{11}^*(\lambda_2) - \mathbf{\Sigma}_{22}^*(\lambda_2)^{-1}
\end{aligned}$$

$$\times \mathbf{\Sigma}_{21}^*(\lambda_2))].$$

$$(A.18)$$

Using the observation equation as in Model (2.11), we have the conditional predictive distribution for y_t^s as follows:

$$
\begin{aligned}
(y_t^s | \mathbf{y}_t, \alpha_{1t}^s, \alpha_{2t}^s, \alpha_{1t}, \alpha_{2t}, \beta_t, \lambda, \sigma^2) \quad \sim \quad & N[\beta_t + S_{1t}(a_1)\alpha_{1t}^s + S_{2t}(a_2)\alpha_{2t}^s + \mathbf{\Sigma}_{12}^*(\lambda) \\
& \times \mathbf{\Sigma}_{22}^*(\lambda)^{-1}(\mathbf{y}_t - \mathbf{1}_n\beta_t - S_{1t}(a_1)\alpha_{1t} \\
& - S_{2t}(a_2)\alpha_{2t}), \sigma^2(\Sigma_{11}^*(\lambda) - \mathbf{\Sigma}_{12}^*(\lambda) \\
& \times \mathbf{\Sigma}_{22}^*(\lambda)^{-1}\mathbf{\Sigma}_{12}^*(\lambda))].
\end{aligned}
$$

$$(A.19)$$

Appendix B

Software for Chapter 3

We write the DLM software, *GDLM.1.0*, using the R and C interface. This software can be freely downloaded from http://enviro.stat.ubc.ca/dlm/ GDLM.1.0.zip or http://enviro.stat.ubc.ca/dlm/GDLM.1.0.tar.gz for windows or unix/linux systems, respectively. The instructions for installing this package in windows and linux/unix is under the folder "INSTALL". This software requires the R software, $R \geq 2.2.0$, and C compiler, that is, MinGW-3.2.0-rc-3.exe. "FFBS" folder which contains the C codes for the forward–filtering–backward–sampling method used in the MCMC algorithm. Their header files are included under the folder "include". "MAIN" folder contains all the R functions for the models used in this chapter. A real database example is written in the folder "DEMO" where the database is located in the folder "DATA". "OUTPUT" allows one to store the computational output from the MCMC algorithm for making some graphical comparisons. The basic information for this software has been summarized in the "Readme" file.

   Two packages in R, *MASS* and *stats*, are required in this software to fit the DLM within Gaussian framework. We now illustrate several important functions in "MAIN" used in this software.

(1) Function "forfun.c" is used as an R interface function with C. It generates two main components to compute the acceptance rate in the Metropolis–within–Gibbs sampling. It has the following arguments: *lat, long, mcmc.data.matrix, lambda, a1, a2, gamma.vec, m.init, C.init* and *BH*. Here are the details for each of these arguments at the jth iteration:

   - "lat" presents the vector of latitudes at the gauged sites.

- "long" presents the vector of longitude at the gauged sites.

- "mcmc.data.matrix" presents the MCMC complete data matrix used at each iteration, that is, $Y^{(j-1)}$. By default, its columns represent responses at each of the time points and rows for each of the sites.

- "lambda" presents the range parameter, λ, in the DLM.

- "a1" presents the phase parameter corresponding to the 24 hour periodicity.

- "a2" presents the phase parameter corresponding to the 12 hour periodicity.

- "gamma.vec" presents fixed hyperparameters, that is, $\gamma = (\tau_y^2, \tau_1^2, \lambda_1, \tau_2^2, \lambda_2)$.

- "m.init" and "C.init" present the initial mean vector and co-variance matrix for the state parameters, that is, $\mathbf{m}_0$ and $\mathbf{C}_0$, respectively.

- "BH" presents the total number of hours included in each day. By default, it is set to be 24.

This function has the following outputs:

- "quad" presents the summation of the quadratic quantities in the log–likelihood of the joint posterior density for all the model parameters, that is, $\sum_{t=1}^T \mathbf{e}_t' \mathbf{Q}_t^{-1} \mathbf{e}_t$.

- "ldet" presents the summation of the log of the determinant quantities in the log–likelihood of the joint posterior density for λ, that is, $\sum_{t=1}^T \log |\mathbf{Q}_t|$.

- "quad.vec" presents the vector of the quadratic quantities calculated at each of the time points, that is, $\{\mathbf{e}_t' \mathbf{Q}_t^{-1} \mathbf{e}_t : t = 1, \ldots, T\}$.

- "ldet.vec" presents the vector of the logarithm of the determinant quantities at each of the time points, that is, $\{\log |\mathbf{Q}_t| : t = 1, \ldots, T\}$.

(2) Function "ffbs.c" is another R interface function with C. It calls a C function in R, "DLMFFBS", to generate the state parameters, $\mathbf{x}_t$, at each MCMC iteration. This function contains the following arguments:

- "lat" presents the vector of latitudes at the gauged sites.

- "long" presents the vector of longitude at the gauged sites.

- "mcmc.data.matrix" presents the mcmc complete data matrix used at each iteration, that is, $\mathbf{Y}^{(j-1)}$. By default, its columns represent responses at each of the time points and rows for each of the sites.

- "lambda" presents the range parameter, λ, in the DLM.

- "sigma" presents the variance parameter, σ^2, in the DLM.

- "a1" presents the phase parameter corresponding to the 24 hour periodicity.

- "a2" presents the phase parameter corresponding to the 12 hour periodicity.

- "gamma.vec" presents fixed hyperparameters, that is, $\gamma = (\tau_y^2, \tau_1^2, \lambda_1, \tau_2^2, \lambda_2)$.

- "m.init" and "C.init" present the initial mean vector and co-variance matrix for the state parameters, that is, $\mathbf{m}_0$ and $\mathbf{C}_0$, respectively.

- "BH" presents the total number of hours included in each day. By default, it is set to be 24.

- "ctr" presents the current index for the iterations in the MCMC algorithm, that is, j in this case.

This function has the following outputs:

- "quad" presents the summation of the quadratic quantities in the log–likelihood of the joint posterior density for all the model parameters, that is, $\sum_{t=1}^{T} \mathbf{e}_t' \mathbf{Q}_t^{-1} \mathbf{e}_t$.

- "ldet" presents the summation of the log of the determinant quantities in the log–likelihood of the joint posterior density for the model parameters, that is, $\sum_{t=1}^{T} \log |\mathbf{Q}_t|$.

(3) Function "GDLM" is an R function to implement the MCMC algorithm for the DLM given by (2.14)–(2.15) in Chapter 2. This function contains the following arguments:

- "srm.data" presents the initial data matrix for the MCMC algorithm where the missing values are fitted by the spatial regression method.

- "origin.data" presents the raw data matrix, including the missing values. By default, its row represents the stations and columns, the time points.

- "job" is a binary variable. "job=0" means this function will only do the MCMC sampling; while "job=1" means this function will do both the MCMC sampling and spatial interpolation at ungauged sites.

- "glat" presents the vector of latitude (in degree) at gauged sites.

- "glong" presents the vector of longitude (in degree) at gauged sites.

- "uglat" presents the vector of latitude (in degree) at the ungauged sites.

- "uglong" presents the vector of longitude (in degree) at ungauged sites.

- "gamma.vec" presents fixed hyperparameters, that is, $\gamma = (\tau_y^2, \tau_1^2, \lambda_1, \tau_2^2, \lambda_2)$.

- "m.init" and "C.init" present the initial mean vector and covariance matrix for the state parameters, that is, $\mathbf{m}_0$ and $\mathbf{C}_0$, respectively.

- "BH" presents the total number of hours included in each day. By default, it is set to be 24.

196

- "ITER" presents the total number of iterations used in the MCMC sampling.

- "tuning.para" presents the fixed tuning parameter, τ^2, for the Metropolis–Hasting step.

- "alpha.init" and "beta.init" present the hyperprior for the variance parameter, σ^2, that is, α_σ and β_σ, respectively.

- "eta.init" and "delta.init" present the hyperprior for the range parameter, λ.

- "mu.phase.init" and "Sigma.phase.init" present the hyperprior for the phase parameters, (a_1, a_2), that is, μ_0 and $\boldsymbol{\Sigma}_0$, respectively.

- "output.direc.name" presents the directory name that the user want to store the output.

This function has the following outputs:

- "lambda" presents the MC samples for the range parameter, λ.

- "sigma" presents the MC samples for the variance parameter, σ^2.

- "phase.a1" and "phase.a2" present the MC samples for the phase parameters, a_1 and a_2, respectively.

- "accept.ratio" presents the acceptance ratio for the range parameter, λ, in the Metropolis–Hasting step.

- "accept.index" presents the index of accepted iteration in the Metropolis–Hasting step.

- "quad" presents the quadratic form at each of the iterations (only $\lambda = \lambda^{(j)}$ and all the other parameters having their $(j-1)^{\text{th}}$ iteration value).

- "log.det" presents the logarithm of the determinant quantity at each of the iterations (only $\lambda = \lambda^{(j)}$ and all the other parameters having their $(j-1)^{\text{th}}$ iteration value).

- "theta.mat" presents the sampling matrix of the state parameters, $\mathbf{x}_{1:\text{ITER}}$, from the MCMC algorithm.

(4) Function "GDLM.INIT" is an R function, having almost the same settings as "GDLM". The difference is that "GDLM.INIT" allows the MCMC sampling to start from any given values for model parameters: λ, σ^2, a_1 and a_2. This function contains most arguments and same output as that of "GDLM" except for the following four arguments in the input:

- "lambda.init.value" presents the starting values for λ.
- "sigma.init.value" presents the starting values for σ^2.
- "phase.a1.init.value" presents the starting values for a_1.
- "phase.a2.init.value" presents the starting values for a_2.

(5) Function "interpolate.fun" is an R function used to interpolate the response variable at ungauged sites. It contains the following arguments:

- "srm.data" presents the initial data matrix for the MCMC algorithm where the missing values are fitted by the spatial regression method.
- "origin.data" presents the raw data matrix, including the missing values. By default, its row represents the stations and columns, the time points.
- "job" is a binary variable. "job=0" means this function will only do the MCMC sampling; while "job=1" means this function will do both the MCMC sampling and spatial interpolation at ungauged sites.
- "glat" presents the vector of latitude (in degree) at gauged sites.
- "glong" presents the vector of longitude (in degree) at gauged sites.
- "uglat" presents the vector of latitude (in degree) at ungauged sites.
- "uglong" presents the vector of longitude (in degree) at ungauged sites.

- "lambda.mcmc" presents the MCMC samples for the range parameter, λ.

- "sigma.mcmc" presents the MCMC samples for the variance parameter, σ^2.

- "phase.a1.mcmc" presents the MCMC samples for the phase parameter, a_1.

- "phase.a2.mcmc" presents the MCMC samples for the phase parameter, a_2.

- "BURN.IN" presents the burn–in period setting for the MCMC samples for the model parameters.

- "m.init" and "C.init" present the initial mean vector and covariance matrix for the state parameters, that is, $\mathbf{m}_0$ and $\mathbf{C}_0$, respectively.

- "gamma.vec" presents fixed hyperparameters, that is, $\gamma = (\tau_y^2, \tau_1^2, \lambda_1, \tau_2^2, \lambda_2)$.

- "BH" presents the total number of hours included in each day. By default, it is set to be 24.

The output of this function is the interpolated values for the response variables at ungauged sites.

Appendix C

Additional Results for Chapter 6

C.1 Additional Results for Section 6.4

Definition C.1.1 *Suppose the random matrix response $\boldsymbol{X} : r \times q$ has a matrix normal distribution, denoted by $\boldsymbol{X} \sim N_{r \times q}(\boldsymbol{M}, \boldsymbol{C}, \boldsymbol{\Sigma})$, where $\boldsymbol{C} : r \times r > 0$, and $\boldsymbol{\Sigma} : q \times q > 0$. Then the probability density function of $\boldsymbol{X}$ is given by*

$$p(\boldsymbol{X}) \;\; = \;\; (2\pi)^{-rq/2} |\boldsymbol{C}|^{-q/2} |\boldsymbol{\Sigma}|^{-r/2} \exp\{-\frac{1}{2} tr[(\boldsymbol{X} - \boldsymbol{M})' \boldsymbol{C}^{-1} (\boldsymbol{X} - \boldsymbol{M}) \boldsymbol{\Sigma}^{-1}]\}. \tag{C.1}$$

Definition C.1.2 *Suppose the random matrix $\boldsymbol{X} : q \times q$ is symmetric, positive definite and follows an inverted Wishart distribution with degrees of freedom δ and scale matrix $\boldsymbol{S}$. Then the probability density function of $\boldsymbol{X}$ is given by*

$$p(\boldsymbol{X}) \;\; = \;\; k |\boldsymbol{X}|^{-(\frac{\delta}{2}+q)} \exp\left\{-\frac{1}{2} tr\left[\boldsymbol{S}\boldsymbol{X}^{-1}\right]\right\}, \tag{C.2}$$

where $\boldsymbol{S}$ is positive definite and

$$k^{-1} \;\; = \;\; 2^{qv/2} \pi^{q(q-1)/4} \prod_{j=1}^{q} \Gamma\left(\frac{v - j - 1}{2}\right) |\boldsymbol{S}|^{-v/2},$$

with $v = \delta + q - 1$.

Proof C.1.1 *(Lemma 6.4.2) By the KL expansion and Lemma 6.4.1, we*

have $\boldsymbol{\Lambda}^{-2} = \boldsymbol{O}\boldsymbol{\Sigma}_S^{-1}\boldsymbol{O}'$ and moreover, $\boldsymbol{\Lambda}^{-2} \sim W_p(n, \boldsymbol{I}_p)$. Hence, the $\{\lambda_j^{-2} : j = 1, \ldots, p\}$ are mutually independent and $\lambda_j^{-2} \sim W_1(n, 1)$, that is, χ_n^2, for $j = 1, \ldots, p$. $\diamond$

Proof C.1.2 *(Theorem 6.4.1) Given* $\boldsymbol{Y} : p \times n \sim N_{p \times n}(\boldsymbol{0}, \boldsymbol{\Sigma}_S \otimes \boldsymbol{\Sigma}_T)$, *denote* $\boldsymbol{Y}^*$ *to be* $\boldsymbol{Y}\boldsymbol{\Sigma}_T^{-1/2}$. *Consequently,* $\boldsymbol{Y}^* \sim N_{p \times n}(\boldsymbol{0}, \boldsymbol{\Sigma}_S \otimes \boldsymbol{I}_n)$. *Similarly, we have* $\boldsymbol{\Sigma}_S^{-1/2}\boldsymbol{Y}^* \sim N_{p \times n}(\boldsymbol{0}, \boldsymbol{I}_p \otimes \boldsymbol{I}_n)$.

By Lemma 6.4.1, $\boldsymbol{\Sigma}_S^{-1/2}\boldsymbol{Y}^* = \boldsymbol{OLP}$, *where* $\boldsymbol{O}$ *represents an orthogonal matrix that is uniformly distributed over the Grassmann manifold,* $\boldsymbol{P}$, *an orthogonal frame that is uniformly distributed over the Stiefel manifold, and* $\boldsymbol{L}$, *a diagonal matrix with entries* $\{l_1, \ldots, l_p\}$ *such that* $l_1^2, \ldots, l_p^2$ *are the eigenvalues for* $(\boldsymbol{\Sigma}_S^{-1/2}\boldsymbol{Y}^*)(\boldsymbol{\Sigma}_S^{-1/2}\boldsymbol{Y}^*)'$. *Hence,* $\boldsymbol{Y}^* = \boldsymbol{\Sigma}_S^{-1/2}\boldsymbol{OLP}$. *Moreover, since*

$$
\begin{aligned}
E[(\boldsymbol{Y}^*)(\boldsymbol{Y}^*)'] &= \boldsymbol{\Sigma}_S^{1/2}E[\boldsymbol{OL}^2\boldsymbol{O}']\boldsymbol{\Sigma}_S^{1/2} \\
&= n\boldsymbol{\Sigma}_S,
\end{aligned}
$$

the Bayesian EOFs can then be given by $\boldsymbol{W} = \frac{1}{n}\boldsymbol{\Sigma}_S^{1/2}\boldsymbol{OL}$. $\diamond$

Proof C.1.3 *(Theorem 6.4.2) By Definition C.1.2, we have*

$$
p(\boldsymbol{\Sigma}_S^{-1}) \quad \propto \quad |\boldsymbol{\Sigma}_S|^{-(\frac{\delta_S}{2}+p)}\exp\{-\frac{1}{2}tr(\boldsymbol{\Xi}_S^{-1}\boldsymbol{\Sigma}_S^{-1})\}.
$$

Given $\boldsymbol{\Sigma}_T$, $\boldsymbol{Y}^* = \boldsymbol{Y}\boldsymbol{\Sigma}_T^{-1/2} \sim N_{p \times n}(\boldsymbol{0}, \boldsymbol{\Sigma}_S \otimes \boldsymbol{I}_n)$. *By Definition C.1.1, we have*

$$
p(\boldsymbol{Y}^*|\boldsymbol{\Sigma}_S) \quad \propto \quad |\boldsymbol{\Sigma}_S|^{-n/2}\exp\{-\frac{1}{2}tr[\boldsymbol{Y}^*(\boldsymbol{Y}^*)'\boldsymbol{\Sigma}_S^{-1}]\}.
$$

Then the posterior distribution for $\boldsymbol{\Sigma}_S^{-1}$ *given* $\boldsymbol{Y}^*$, *that is,* $\boldsymbol{Y}$ *for known* $\boldsymbol{\Sigma}_T$ *is given as follows:*

$$
\begin{aligned}
p(\boldsymbol{\Sigma}_S^{-1}|\boldsymbol{Y}) &\propto p(\boldsymbol{Y}^*|\boldsymbol{\Sigma}_S)p(\boldsymbol{\Sigma}_S^{-1}) \\
&\propto |\boldsymbol{\Sigma}_S|^{-(\frac{\delta_S+n}{2}+p)}\exp\left\{-\frac{1}{2}tr\left[(\boldsymbol{\Xi}_S^{-1} + \boldsymbol{Y}^*(\boldsymbol{Y}^*)')\boldsymbol{\Sigma}_S^{-1}\right]\right\}.
\end{aligned}
$$

In other words, $\mathbf{\Sigma}_S^{-1}|\, \mathbf{Y} \sim W_p(\delta^o, \mathbf{\Xi}^o)$, where $\delta^o = \delta_S + n$ and

$$
\begin{aligned}
\mathbf{\Xi}^o &= \{\mathbf{\Xi}_S^{-1} + \mathbf{Y}^*(\mathbf{Y}^*)'\}^{-1} \\
&= (\mathbf{\Xi}_S^{-1} + \mathbf{Y}\mathbf{\Sigma}_T^{-1}\mathbf{Y}')^{-1} \\
&= \mathbf{\Xi}_S - \mathbf{\Xi}_S\,\mathbf{Y}(\mathbf{Y}'\mathbf{\Xi}_S\,\mathbf{Y} + \mathbf{\Sigma}_T)^{-1}\,\mathbf{Y}'\mathbf{\Xi}_S.
\end{aligned}
$$

$\diamond$

Proof C.1.4 *(Theorem 6.4.3) This theorem can be proved similarly as in Proof C.1.3, and so omitted here.* $\diamond$

Proof C.1.5 *(Theorem 6.4.4) Let $\mathbf{V} = \mathbf{\Sigma}_S^{-1/2}\mathbf{Y}$. Then $\mathbf{V} \sim N_{p \times n}(\mathbf{0}, \mathbf{I}_p \otimes \mathbf{\Sigma}_T)$. Hence, we have*

$$
p(\mathbf{Y}|\mathbf{\Sigma}_S, \mu, \theta) \;\propto\; \exp\left\{-\frac{1}{2}tr\left[\mathbf{V}'\mathbf{V}\rho(.,\theta)^{-1}\right]\right\}.
$$

Given the prior for θ, $N_k(\theta, \mathbf{\Sigma}_0)$, the posterior conditional density for θ can be represented by

$$
\begin{aligned}
p(\theta|\,\mathbf{Y}, \mathbf{\Sigma}_S, \mu) &\propto\; p(\mathbf{Y}|\mathbf{\Sigma}_S, \mu, \theta)p(\theta) \\
&\propto\; \exp\left\{-\frac{1}{2}\left[tr(\mathbf{V}\mathbf{V}'\rho(.,\theta)^{-1}) + (\theta - \theta_0)'\mathbf{\Sigma}_0^{-1}(\theta - \theta_0)\right]\right\}.
\end{aligned}
$$

$\diamond$

Proof C.1.6 *(Theorem 6.4.5)*

(i) Since

$$
\mathbf{Z} \sim N_{p \times n}(\mu \otimes \mathbf{1}_n', \mathbf{\Sigma}_S \otimes \mathbf{\Sigma}_T),
$$

and $p(\mu) \propto 1$, we have the posterior conditional distribution for μ as follows:

$$
\begin{aligned}
p(\mu|\mathbf{Z}, \mathbf{\Sigma}_S, \mathbf{\Sigma}_T) &\propto\; p(\mathbf{Z}|\mu, \mathbf{\Sigma}_S, \mathbf{\Sigma}_T)p(\mu) \\
&\propto\; \exp\{-\frac{1}{2}tr[(\mu \otimes \mathbf{1}_n' - \mathbf{Z})\mathbf{\Sigma}_T^{-1}(\mu \otimes \mathbf{1}_n' - \mathbf{Z})'\mathbf{\Sigma}_S^{-1}]\} \\
&\propto\; \exp\{-\frac{1}{2}tr[(\mu \otimes \mathbf{1}_n')\mathbf{\Sigma}_T^{-1}(\mu \otimes \mathbf{1}_n')'\mathbf{\Sigma}_S^{-1} - (\mu \otimes \mathbf{1}_n')
\end{aligned}
$$

$$\times \mathbf{\Sigma}_T^{-1} \mathbf{Z}' \mathbf{\Sigma}_S^{-1} - \mathbf{Z}\mathbf{\Sigma}_T^{-1}(\mu \otimes \mathbf{1}_n')'\mathbf{\Sigma}_S^{-1}]\}$$

$$= \exp\{-\frac{1}{2}tr[(\mu\mu' tr(\ \mathbf{1}_n'\mathbf{1}_n\mathbf{\Sigma}_T^{-1}) - \mu\mathbf{1}_n\mathbf{\Sigma}_T^{-1}\mathbf{Z}'$$

$$-\mathbf{Z}\mathbf{\Sigma}_T^{-1}\mathbf{1}_n'\mu')\mathbf{\Sigma}_S^{-1}]\}$$

$$\propto \exp\{-\frac{1}{2}tr[(\mu - \mathbf{M})(\Sigma^*)^{-1}(\mu - \mathbf{M})'\mathbf{\Sigma}_S^{-1}]\},$$

where $\Sigma^* = \{tr(\mathbf{1}_n'\mathbf{1}_n\mathbf{\Sigma}_T^{-1})\ \}^{-1}$ and $\mathbf{M} = \mathbf{Z}\mathbf{\Sigma}_T^{-1}\mathbf{1}_n'\Sigma^*$. Therefore, we have

$$\mu|\mathbf{Z}, \mathbf{\Sigma}_S, \mathbf{\Sigma}_T \quad \sim \quad N_{1 \times p}(\mathbf{M}, \Sigma^* \otimes \mathbf{\Sigma}_S),$$

that is, $N_p(\mathbf{M}, \Sigma^*\mathbf{\Sigma}_S)$ since Σ^* is a scalar.

(ii) Let $\mathbf{Y} = \mathbf{Z} - \mu \otimes \mathbf{1}_n'$. Consequently, we have $\mathbf{Y}\mathbf{\Sigma}_T^{-1/2} \sim N_{p \times n}(\mathbf{0}, \mathbf{\Sigma}_S \otimes \mathbf{I}_n)$. By Theorem 6.4.2, the posterior distribution for $\mathbf{\Sigma}_S^{-1}|\mathbf{Z}, \mathbf{\Sigma}_T$ is $W_p(\delta^o, \mathbf{\Xi}^o)$, given by (6.16), where $\mathbf{Y} = \mathbf{Z} - \mu \otimes \mathbf{1}_n'$.

(iii) Similarly as in (ii), by Theorem 6.4.3 the posterior conditional distribution for $\mathbf{\Sigma}_T^{-1}|\mathbf{Z}, \mathbf{\Sigma}_S$ is given by (6.18), where $\mathbf{Y} = \mathbf{Z} - \mu \otimes \mathbf{1}_n'$.

$\diamond$

Proof C.1.7 *(Theorem 6.4.6) The proof for this theorem follows Theorem 6.4.4 but letting $\mathbf{V} = \mathbf{\Sigma}_S^{-1/2}(\mathbf{Z} - \mu \otimes \mathbf{1}_n')$.* $\diamond$