

STOCHASTIC MODELING OF SPACE - TIME PROCESSES.  
AN AIR POLLUTION PROBLEM

by

STOITCHKO KALENDERSKI

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDENTS  
(Atmospheric Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA  
(Vancouver)

April 2009

© Stoitchko Kalenderski, 2009

## **ABSTRACT**

This study presents an interdisciplinary approach to an air pollution problem that takes into account the physical bases that govern the processes of interest under the framework of a Bayesian spatiotemporal model. Based on this approach we have developed a physically motivated stochastic model that decomposes the ground-level pollutant concentration field in three components, namely: transport, local production, and large-scale mean trend mostly dominated by the emission rates. The model highlighted the importance of simultaneously considering different aspects of an air pollution problem. This approach is novel in the field of environmental spatial statistics and gives a different perspective on the subject.

Based on the advection equation we have modeled the transport component. The advection equation implies that two important factors have to be considered when modeling the transport of an air pollutant namely: wind field and the pollutant gradients. The equation also shows that these two factors should be considered concurrently.

The local production component ignores contributions from neighboring sites and takes into account only specific local meteorological conditions. This approach reflects the fact that physics and chemistry do not depend on location and time. As a result the local production spatial structure is not modeled directly, but is inherent in the covariate fields.

To specify the large-scale mean process, which is mostly dominated by the emissions rates, we used the first eigenfunction of a Principal Component Analysis performed on wind pre-filtered dataset. This new approach allowed us to capture very complex spatial patterns and to avoid using emissions inventory data. The emissions rates reveal the different capacities of a pollutant production throughout the domain of interest.

The proposed models were applied to a Lower Fraser Valley dataset and were able to quantify the transport and local contributions to ozone fields. The models also capture a number of key features of ozone behavior and show excellent agreement with the observed data and outperformed considerably simpler univariate time series models by a factor of 2 to 3.

# TABLE OF CONTENTS

ABSTRACT .....	ii
TABLE OF CONTENTS .....	iv
LIST OF TABLES .....	vii
LIST OF FIGURES .....	ix
NOTATIONS, ABBREVIATIONS AND CONVENTIONS.....	xi
ACKNOWLEDGEMENTS .....	xiii
DEDICATION .....	xv
<b>1 INTRODUCTION .....</b>	<b>1</b>
1.1 Modeling an air pollution problem .....	1
1.2 Research problems, motivation and objectives .....	3
1.2.1 Research problems .....	3
1.2.2 Motivation .....	3
1.2.3 Research objectives .....	5
1.3 Organization of dissertation .....	5
<b>2 MODEL DEVELOPMENT .....</b>	<b>7</b>
2.1 Statistical methods for spatial-temporal data .....	7

2.1.1	Regression models .....	7
2.1.2	Extreme value approaches .....	8
2.1.3	Space-time models .....	9
2.2	Model requirements and modeling strategy .....	11
2.3	A hierarchical spatiotemporal dynamic model .....	12
2.3.1	Data model .....	12
2.3.2	Process model .....	13
2.3.2.1	Large-scale spatiotemporal trends .....	14
2.3.2.2	Transport model .....	16
2.3.2.3	Local production model .....	20
2.3.2	Parameters model .....	22
2.4	Prediction.....	23
3	<b>SIMULATION STUDY .....</b>	<b>25</b>
3.1	Boundary sites .....	25
3.2	The autoregressive process.....	28
3.3	Spatial data coverage.....	29
3.4	Prior information .....	31
3.3	Results .....	32
3.5.1	Parameters inference .....	33

3.5.2	Model performance .....	42
3.3	Discussion .....	50
<b>4</b>	<b>APPLICATION TO GROUND LEVEL OZONE CONCENTRA- TION DATA .....</b>	<b>52</b>
4.1	The Metro Vancouver data .....	52
4.2	Background chemistry of ozone .....	55
4.3	Exploratory Data Analysis.....	57
4.4	Results .....	63
4.4.1	Parameter inference .....	63
4.4.2	Model performance .....	67
4.4.3	Model comparison .....	75
4.5	Discussion .....	77
<b>5</b>	<b>CONCLUSIONS AND FUTURE RESEARCH .....</b>	<b>79</b>
	<b>REFERENCES .....</b>	<b>82</b>
	<b>APPENDIX A .....</b>	<b>89</b>
	<b>APPENDIX B .....</b>	<b>100</b>

## LIST OF TABLES

<b>Table 3.1</b>	Summary of the noninformative prior specification.....	31
<b>Table 3.2</b>	Summary of the informative prior specification.....	32
<b>Table 3.3</b>	Summary of the posterior parameter inference with different number $n$ of available observations based on model $Model_1$ and noninformative prior (Table 3.1) .....	34
<b>Table 3.4</b>	Summary of posterior parameter inference base on model $Model_1$ using different number $n$ of available observations and informative prior (Table 3.1).....	35
<b>Table 3.5</b>	Summary of posterior parameter inference base on model $Model_2$ , using different number $n$ of available observations and noninformative prior (Table 3.1).....	38
<b>Table 3.6</b>	Summary of posterior parameter inference base on model $Model_2$ , using different number $n$ of available observations and informative prior (Table 3.2).....	38
<b>Table 3.7</b>	SD statistics for spatial varying autoregressive process $\rho$ .....	38
<b>Table 3.8</b>	Summary of posterior parameter inference base on model $Model_3$ , using different number $n$ of available observations and noninformative prior (Table 3.1).....	39
<b>Table 3.9</b>	Summary of posterior parameter inference base on model $Model_3$ , using different number $n$ of available observations and informative prior (Table 3.2).....	40
<b>Table 3.10</b>	Summary of posterior parameter inference base on model $Model_4$ , using different number $n$ of available observations and noninformative prior (Table 3.1).....	40
<b>Table 3.11</b>	Summary of posterior parameter inference base on model $Model_4$ , using different number $n$ of available observations and informative prior (Table 3.2).....	41
<b>Table 3.12</b>	Summary of the models performance and relative importance of the transport effect. * denotes that informative prior was used for calculations.....	43
<b>Table 4.1</b>	Summary of posterior parameter inference .....	64
<b>Table 4.2</b>	Estimation of boundary parameter.....	64
<b>Table 4.3</b>	Summary of the overall models performance and the relative importance of the transport effect .....	67
<b>Table 4.4</b>	Comparison of model performance [ppb].....	76
<b>Table B.1</b>	Ozone summaries for three years studied.....	99
<b>Table B.2</b>	Meteorological data summaries for three years studied .....	99
<b>Table B.3</b>	Model $Model_1$ parameter and performance statistics summaries for three years studied .....	100

<b>Table B.4</b>	Model <i>Model<sub>2</sub></i> parameter and performance statistics summaries for three years studied .....	101
<b>Table B.5</b>	Model <i>Model<sub>3</sub></i> parameter and performance statistics summaries for three years studied .....	102
<b>Table B.6</b>	Model <i>Model<sub>4</sub></i> parameter and performance statistics summaries for three years studied .....	103



## LIST OF FIGURES

<b>Figure 3.1</b>	Selected grid cells for the four datasets. Cells are counted from the left bottom corner .....	30
<b>Figure 3.2</b>	Sequences of 5000 iterations from the posterior distributions of the model $Model_1$ parameters and the histograms of the last 2000 iterations.....	36
<b>Figure 3.3</b>	True and interpolated surfaces of $\rho$ based on informative prior and $n = 32$ . Shading interval is uniform .....	39
<b>Figure 3.4</b>	Left: sequences of 5000 iterations from the posterior distributions of the randomly selected grid cells without observations $O(9,20)$ and with observations $O(25,20)$ ; Center: histogram of the last 2000 iterations; Right: autocorrelation function.....	44
<b>Figure 3.5</b>	Posterior mean time series for grid cells without observations $O(9,\cdot)$ and for grid cells with observation $O(25,\cdot)$ .....	45
<b>Figure 3.6</b>	Model-estimated mean field of the true pollutant concentration (grey scale), and observed values (numbers). * - denotes the grid cells used for fitting the model .....	46
<b>Figure 3.7</b>	Contour plots of pollutant variance field. * - denotes the grid cells used for fitting the model.....	47
<b>Figure 3.8</b>	Model estimated mean transport contributions map. . The observed wind field is superimposed on the plots .....	48
<b>Figure 3.9</b>	Scatter plot for grid cell with observations $O(25,\cdot)$ and for grid cell without observation $O(2,\cdot)$ .....	49
<b>Figure 4.1</b>	Interpolated 8-hour average wind field for 12 May 2004.....	53
<b>Figure 4.2</b>	Locations of ozone observation sites “*” and prediction grid “+”. Numbers are station identifiers.....	54
<b>Figure 4.3</b>	Boxplot of the domain daily maximum 8-hour ozone levels by month .....	57
<b>Figure 4.4</b>	Time series of observed daily maximum 8-hour zone concentrations at 20 monitoring sites from April 15th to September 30th.....	58
<b>Figure 4.5</b>	The first EOF modes of ozone data .....	60
<b>Figure 4.6</b>	Normal Q-Q plot of residuals on original scale (top panels), square root scale (middle panels), and logarithmic scale (bottom panels).....	62
<b>Figure 4.7</b>	Partial autocorrelation functions of the first nine time series .....	62
<b>Figure 4.8</b>	Estimated surfaces of autoregressive process $\rho(\cdot)$ for models $Model_2$ and $Model_3$ .....	64
<b>Figure 4.9</b>	Sequences of 5000 iterations from the posterior distributions of the model $Model_2$ parameters and the histograms of the last 2000 iterations.....	66

<b>Figure 4.10</b>	Left: sequences of 5000 iterations from the posterior distributions of the randomly selected grid cells without observations O(125,20) and with observations O(71,20) ; Center: histogram of the last 2000 iterations; Right: autocorrelation function.....	68
<b>Figure 4.11</b>	Time series for cells (75) with observations (station 30).....	69
<b>Figure 4.12</b>	Model-estimated mean field of the true pollutant concentration as residuals. The observed values are super imposed on the plots .....	71
<b>Figure 4.13</b>	Contour plots of pollutant variance field. * - denotes the monitoring sites locations .....	72
<b>Figure 4.14</b>	Model estimated mean transport contributions map. The observed wind field is superimposed on the plots .....	73
<b>Figure 4.15</b>	Scatterplot of observed ( $x$ -axis) and modeled ( $y$ -axis) concentrations for models $Model_1$ , $Model_2$ , $Model_3$ , and $Model_4$ on the scale of residuals.....	74

## NOTATIONS, ABBREVIATIONS AND CONVENTIONS

Whenever possible, we follow the widely use conventions throughout the thesis. A random variable or a matrix are denoted by upper-case Roman letter e.g.  $X$ . Bold upper-case Roman letter are used for random vectors e.g.  $\mathbf{X}$ . Components of a vector  $\mathbf{X}$  are denoted by  $(X_1, X_2 \dots)$  whereas components of sequence will tend to be denoted by  $x^{(1)}, x^{(2)}, \dots$ .

The vector  $\mathbf{X}_{-i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$  is the vector  $\mathbf{X}$  with its  $i$ th component removed,  $i = 1, \dots, n$ . Greek letters are used to denote parameters and parameter vectors e.g.  $\theta, \boldsymbol{\theta}$ . For a random vector  $\mathbf{X}$ ,  $[\mathbf{X}]$  represent joint probability density function and  $[\mathbf{X}|\mathbf{Y}]$  represent conditional density of  $\mathbf{X}$  given the vector  $\mathbf{Y}$ .

$\sim$	distributed as
$\approx$	almost equal to
$\propto$	proportional to
$\otimes$	Kronecker product
$I$	Identity matrix
$\mathbf{1}_n$	$n$ vector of ones
$\mathbf{0}_n$	$n$ vector of zeros
$diag(\mathbf{a})$	diagonal matrix with elements of the vector $\mathbf{a}$ on the main diagonal
$N(\mu, \sigma^2)$	univariate normal distribution with mean $\mu$ , and variance $\sigma^2$
$N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$p$ -variate Multivariate Normal distribution with mean vector $\boldsymbol{\mu}$ and cova-

	riance matrix $\Sigma$ .
process	a physical variable of interest
$IG(\alpha, \beta)$	univariate inverse gamma distribution with shape $\alpha$ and rate $\beta$
MRF	Markov random field
$\hat{\cdot}$	Estimated values
$\sim$	Prior values

## ACKNOWLEDGEMENTS

First of all, I would like to thank my two research supervisors, Dr. D.G. Steyn and Dr. J.V. Zidek for their encouragement, guidance and intellectual support. Their insightful suggestions and comments greatly helped me to clarify my thoughts and extend the research in many aspects. I especially thank Dr. Steyn for helping on several times with revisions of this thesis and helping me to master basic writing skills in English. I am grateful to him for giving me the opportunity to participate in Greenapple research project, and for funding my research from his grants. I am also grateful to Dr. Zidek for his kindness and for giving me the opportunity to participate in the PIMS summer school, University of Washington July 9-13, 2007.

I am pleased to thank the other two members of my committee, Dr. P. Austin and Dr. S. Galmarini for their help and encouragements on each stage of my research and especially for their constructive comments on the final draft of this thesis. Their help and support is greatly appreciated. Additionally, I gratefully acknowledge the MatLab-code provided by Dr. N.J. McMillan from the Battelle Memorial Institute, USA, which helped me to organize the implementation of the Markov Chain Monte Carlo algorithm used in this study.

I am grateful to the Metro Vancouver Air Quality Division for providing me (through an agreement with my research supervisor, and assisted by Al Percival) their entire air pollutant and meteorological data set covering the 24 year period from 1982 to 2006. Without these data, my research would not have been possible.

I also wish to acknowledge my friends and fellow students, who have accompanied me throughout these four years at the university. I particularly want to thank Dr.

Bruce Ainslie and Dr. Christian Reuter for their help, support and the many meaningful discussions we have during these pas years.

Finally, I would like to thank my family who always supported me unconditional-ly. Their encouragement and love gave me strength and confidence which helped me to overcome many obstacles and to get the job done.

## DEDICATION

*For my father,  
who would be so proud.*

# CHAPTER 1

## INTRODUCTION

In studies of air pollution, a central question is the spatial prediction and forecasting of pollutant concentration. We need to know pollutant concentrations because in general the severity of the effect depends on the concentration level of the pollutant. Because direct measurements are costly and difficult, modeling is an often-used approach to determining concentration levels. There are several approaches for modeling an air pollution problem. In the following section a brief review of these approaches is given.

### 1.1 Modeling an air pollution problem

Air pollution models are generally categorized into two types depending on the methods used (Berliner, 2003), namely: deterministic (such as transport dispersion and photo-chemical), and statistical (such as regression and kriging).

Deterministic air pollution models are based on physical and/or chemical relations between the pollutant concentration and other factors that characterize the atmosphere. While these models are fully deterministic in the sense that they solve the equations of physics and chemistry (primarily the Navier-Stokes and continuity equations), they also contain elements that are implicitly statistical since the deterministic relations used in these models are spatially and temporally filtered so that they apply to grid-averaged quantities. We only want to know the average state of the atmosphere. It is neither feasible nor desirable to consider in detail all of the small scale motions (turbulence) that are



present in the atmosphere (Randall, 2006). The averaged deterministic equations are mathematically unclosed without further approximations, which is why the deterministic models also explicitly use statistical methods, for example in the way we model turbulence and diffusion at subfilter scales. Fully deterministic approaches involve a large number of equations and parameters. By employing a statistical approach we are able to greatly reduce the size and complexity of the problem. What we lose in statistical models is the ability to resolve fields in time and space at scales below those defined by the measurement grid. What we gain in deterministic models is the ability to examine physical and chemical processes in detail. Mixed deterministic-statistical approaches retain advantages of both deterministic and statistical approaches.

It should also be noted that deterministic models are initialized by emissions and meteorological data and can be used to predict concentrations at selected locations on a computational grid. The computational grid can be either Eulerian or Lagrangian but a hybrid approach is also possible. Additionally, the implementation of the chemistry in these models introduces significant uncertainty because of deeply parameterized chemistry. Jimenez et al. (2003) provide brief descriptions of the main chemical mechanisms used in air quality models.

A few examples of combined meteorological and chemical models are: MAQSIP – Multiscale Air Quality Simulation Platform (Odman and Ingram, 1996); CMAQ – Community Multiscale Model for Air Quality (Byun et al., 1998); EMEP – European Monitoring and Evaluation Program (Simpson, 1992); EURAD – European Air Dispersion Model (Hass, 1991). The list given here is far from exhaustive: see Russell and Dennis, 2000; Holmes and Morawska, 2006, for a review.

In a situation where scientific understanding of physical and chemical processes is incomplete, statistical models are often more appropriate than deterministic models. The statistical models are not as detailed as the deterministic models but can be more efficient and are still able to address very important questions. Furthermore, statistical approaches

have one important advantage over (deterministic) dynamic forecasting methods: their capacity to produce probabilistic forecasts and explicitly account for uncertainty in general. Its ability to evaluate uncertainty, which is inevitable in complex systems, is an extra desirable characteristic of such model. A concise overview of statistical methods relevant to environmental problems is given in the next chapter.

In my study I intend to adopt a middle-ground modeling approach, one that is intermediate between purely deterministic and purely stochastic. The approach I will take is essentially stochastic but is able to easily incorporate physical and chemical reasoning into the statistical analysis. The spectrum of incorporation of physics could range from a qualitative use of physical arguments to strong reliance on numerical models. While this approach is not new, the review of Berliner (2003) shows that initial applications have been quite preliminary. The present work is a substantial elaboration and extension of the approach.

## **1.2 Research problems, motivation and objectives**

### **1.2.1 Research problems**

In addition to the evaluation of concentration levels, another central question of air pollution, which has not been considered in the context of space-time statistical modeling is *the quantification of the relative importance of regional transport and local production of an air pollutant*. In my dissertation I want to address this question simultaneously with the other two major problems of air pollution: predicting pollutant levels and estimating the associated uncertainty. Obtaining a quantitative estimate of the transport and local contributions together with the prediction of the concentration levels and evaluating un-

certainty will be the main objectives of a stochastic space-time model I will develop below.

### **1.2.2 Motivation**

In many studies, the quantitative estimate of the relative importance of pollutant transport and local production has been of great help in finding solutions to an air pollution problem. Such studies often lead to practical methods for solving air pollution problems once scientific understanding of the nature of the problem has been achieved. Sometimes intuitive approaches can be very misleading if, for example, the analysis is conducted at a large scale while there are smaller scale effects (e.g. of topography or small water bodies) which are important for local weather systems but are not considered in the analysis. Alternatively, local solutions are often put in place while the problem starts thousands of kilometers away. The quantitative estimate of a pollutant transport and local production is crucial for effective control strategies. For example, a small local contribution relative to the transport suggests that risk management plan based on local emission control policies will not likely be effective. I believe that a large part of the reason why atmospheric scientists have not investigated these problems in the context of the spatiotemporal modeling approach is their unfamiliarity with the capabilities this approach provides. Statistical analysis that accounts for both spatial and temporal dependence has seldom been addressed in the literature (Le et al, 2001; Guttorp et al, 1994; Cressie and Huang, 1999). However, these studies usually do not incorporate meteorological variables like wind speed and direction, temperature, solar radiation and pressure into the spatio-temporal covariance structure (Huang and Hsu, 2004).

In addition, statistical approaches to this problem do not require preceding knowledge about the emission rates of anthropogenic and natural precursors for the pollutant, which are necessary if we try to address the problem using deterministic models. Requir-

ing such information could be problematic. Emission rates are not always available; it is generally very expensive to collect and update information about them; and finally they are not constant and are subject to change in time and space (Yin et al., 2004).

An important motivation for the present investigation is the fact that uncertainty, which is an intrinsic element of any modeling approaches and atmospheric system is seldom addressed. Quantification of uncertainty is needed to mitigate the risk of our ignorance and to evaluate the degree of reliability of a forecast. Since the atmosphere is a dynamically chaotic system (Lorenz, 1963) and always incompletely observed, uncertainty is a central issue here at each level of the analysis.

Finally, sometimes it is necessary to predict pollutant concentration levels at locations where there are no monitoring stations. This is because very often locations of measuring sites are selected based on criteria that have very little to do with science (see Wikle for relevant discussion, 1996) in the first place and second because of the effects of constantly increasing urbanization on very different aspects of the ecological system. The class of statistical models we are considering provides estimates of the pollutant levels throughout the domain of interest at each particular point in time.

### **1.2.3 Research objectives**

- Developing a non-stationary space-time model for forecasting ground-levels concentrations of an air pollutant at local to regional scales.
- Quantitative assessment of transport and local production of an air pollutant.
- Understanding and estimations of inherent uncertainty as well as model uncertainty and ultimately uncertainty in model predictions.

### **1.3 Organization of dissertation**

This dissertation is composed of five chapters. Chapter 2 is the most theoretical part of the dissertation. It summarizes the statistical methods used for modeling spatiotemporal data, specifies the model requirements and modeling strategy, and formulates a general hierarchical model for describing environmental datasets. In Chapter 3 we analyze a simulated dataset and investigate how parameter inference, model performance, and transport effect depend on: the way we treat the domain boundary sites; whether we allow the time autoregressive process to vary spatially or not; spatial data availability; and prior information. Chapter 4 shows how to apply the proposed general model to ground-level ozone data. This chapter also provides background chemistry for ozone and extended exploratory data analysis (EAD). A general discussion and future issues to explore are given in Chapter 5.

## **CHAPTER 2**

### **MODEL DEVELOPMENT**

#### **2.1 Statistical Methods for spatial-temporal data**

This section is a brief overview of the statistical methods commonly used in many air pollution studies. We need statistical methods generally to acknowledge our ignorance (and possibly uncertainty) and draw conclusions regarding underlying physical and sampling processes (see Le and Zidek p.53, 2006, for more reasoning and examples). In general, these methods can be grouped in three broad categories: regression modeling, extreme value approaches, and space-time models (Thompson et al., 2001). Within each category there are further distinctions depending for example on: the way meteorological data are introduced (directly or via dimension reduction), whether parametric or non-parametric techniques are used, and the way space-time dynamics are incorporated.

##### **2.1.1 Regression models**

Regression based models are the most frequently used statistical models in air pollution literature, with widely varying degrees of complexity. These models, in turn, can be further, grouped in three major classes: linear regression, regression tree, and non-linear regression. Linear regression models are the simplest form of regression models but have limited flexibility to capture interactions and non-linearities between meteorological and pollutant quantities. Linear models have been used for forecasting surface ozone concen-

tration as well as other pollutants by: Gardner and Dorling (2000), Robeson and Steyn (1989), and McKendry (2002). Classification and Regression Trees (Breiman et al., 1984) are able to represent non-linear relationship between meteorological variables and pollutant concentrations. Regression trees have been used to model daily maximum ozone concentrations (Burrows et al., 1995). However, the trend component cannot be captured directly in such models. Finally, non-linear regression can be appropriate when a non-linear relationship is required by the nature of the physical problem. Cobourn and Hubbard (1999) develop a nonlinear regression model for predicting daily maximum 1-h ozone concentrations. The terms in the model were developed from exploratory analysis using graphical techniques and nonparametric regression to discover the salient relationships involving ozone and many candidate predictor variables.

### **1.2.2 Extreme value approaches**

Due to the inherent averaging in, regression models, they are often poor predictors of extreme values (NRC, 1991). An alternative approach for modeling rare events is to use extreme value theory (Gumbel, 1953). Extreme value approaches are particularly useful in the context of modeling threshold exceedances. One advantage of the extreme value approach is that one can easily compute a forecast probability of exceedance for a day, given its (forecasted) meteorology and current-day maximum ozone. This enables local governments to release ozone alert forecasts (Thompson et al., 2001). Tobías and Scotto (2005) were able to model ozone exceedances of the thresholds established by the air quality guidelines and obtain  $n$  year return level (i.e. the level which is exceeded on average once in  $n$  years), using a Generalized Pareto Distribution (GPD). This model was carried out by considering a single long time-series without relating ozone concentration to meteorological condition. Cox and Chu (1996) developed an extreme value model that

takes into account ozone-meteorology relationship. The proposed model is based on Weibull hazard model applied to an ozone value.

### **2.1.3 Space-time models**

In atmospheric sciences, the most frequently used methods for analyzing space-time processes are the “eigentechniques” such as Empirical Orthogonal Function/Principal Component Analysis (EOF/PCA: Lorenz, 1956), Principle Oscillation Pattern (POP), Singular Spectrum Analysis (SSA), and Canonical Correlation Analysis (CCA). See von Storch and Zwiers (1999) for an overview of these methods. Recent developments in neural network modeling have further led to the nonlinear generalization of these classical methods NLPCA, NLCCA and NLSSA (Hsieh, 2004). These dimension-reduction linear techniques are able to extract dominant patterns of variability from a stochastic process. However, when applying these methods, there are several concerns that deserve particular attention. First, as with any data reduction technique they focus but also filter, and the filtering is what makes this approach incomplete at best and misleading at worst. In some cases the reduced (filtered) dataset may not contain important patterns for the analysis which originally were presented in the data and so leading to wrong conclusions. Second, it is very important that observations on an irregular grid are representative of the domain of interest (Karl et al., 1982). Third, singular value decomposition methods (“eigentechniques”) can be used for predictive purposes (Fuentes et al., 2006).

Another approach, still within the framework of space-time models, which focus on the prediction in space and time is dynamical space-time models. These models are built on the extension of established spatial (Cressie, 1993; Wike and Cressie, 1999) or temporal (Box and Jenkins, 1976) methods such as: kriging and autoregressive models. The dynamical space-time models treat jointly space and time variations and typically decompose the random field of interest into a mean component modeling the trend



(smooth variability) and a residual component modeling the higher frequency fluctuations around that trend. An alternative approach is to decompose the random field into collection of spatially correlated time series (i.e. a multivariate time series model, Bennet, 1979) or collection of temporally correlated space patterns (Simard and Marcotte, 1993). These two approaches can be considered as a two-stage method where the results from the conditional analyses in the first stage are combined in the second stage. This approach is appealing because of its simplicity but do not takes into account the interaction between the spatial and temporal component.

The dynamics (causality) in a space-time model is incorporated directly or indirectly. The direct modeling of causality requires specifying a valid non-separable covariance function. There are several methods for constructing non-separable covariance function for spatio-temporal data some of which are in time domain (Gneiting, 2002) and the other in spectral domain (Cressie and Huang, 1999; Fuentes et al., 2008). See Schabenberger and Gotway p.435, 2005, for an overview. The indirect modeling of causality requires specifying an autoregressive model, that is, specifying lagged relationships in time and space. In the time series context (Box and Jenkins, 1976) ARIMA models are very popular. The autoregressive models are also widespread in the spatial data context, popular approaches include Simultaneous Autoregressive (SAR) Models (Whittle, 1954) and Conditional Autoregressive (CAR) models (Besag, 1974). The above autoregressive models can be generalize for space-time data to space-time autoregressive-moving-average (STARMA) models (Cressie p.449, 1993).

The most widely known and used subclass of dynamic models are the Dynamic Linear Models (DLM), or state-space models (West and Harrison, 1997). A DLM framework is a common approach used in environmental problems, which allows updating uncertainties about model parameters as observations are made. The Kalman filter and state-space smoothing algorithms are used to carry out the computations (Carter and Kohn, 1994). The DLM can be seen as a generalization of the regression models that al-

lows time varying coefficients. Recently, the DLM has been applied successfully to multidimensional (space) time series including: Huerta et al., 2004; Stroud et al., 2001.

Recently, several dynamic spatiotemporal models have been reported in literature with application to air pollution problems: Sahu et al., 2007; McMillan et al., 2005; Huang and Hsu, 2004; Huerta et al., 2004; Wikle and Cressie, 1999.

## **2.2 Model requirements and modeling strategy**

To incorporate the physical nature of environmental processes and to meet the objectives specified in Chapter 1, we require development of statistical methods that take into account:

- Dynamical nature of the modeled processes
- Interaction between the spatial and temporal components
- Dependence on other physical processes
- Available deterministic relationships
- Uncertainties

These complexities also require a sophisticated modeling strategy. The Bayesian hierarchical modeling paradigm provides a natural framework for achieving these goals (Berliner, 2003).

The Bayesian Hierarchical Modeling (BHM) approach has several advantages which make it very attractive. First, it allows separation of *statistical* modeling of observations from *stochastic* modeling of the underlying physical processes. Second, since the outputs of a Bayesian model are probability distributions, the approach focuses on uncertainty analysis. Third, prior information concerning previous experience or scientific understanding can be easily incorporated into the analysis. The essence of BHM is con-

structuring a high-dimensional joint probability distribution for all unknowns as the product of comparatively simple conditional probability models. Developing BHM usually goes through organizing variables in three levels (Wikle et. al, 1998):

*Data model:* distribution of observations, conditional on process and model parameters.

*Process model:* distribution of process of interest, conditional on parameters.

*Parameter model:* prior distribution of all model parameters.

Then BHM takes the following form:

$$[process, parameters|data] \propto [data|process, parameters] \times [process|parameters] \times [parameters]$$

where the bracket notation  $[.]$  represents a probability distribution.

## 2.3 A hierarchical spatiotemporal dynamic model

### 2.3.1 Data model

At this first stage of the hierarchy we usually assume that the data are conditionally independent which leads to relatively easy construction of the likelihood function and simplified calculations. The general model considered here is for spatiotemporal data recorded at  $n$  locations  $s_i, i = 1, \dots, n$ , in a gridded spatial domain  $D \subset R^2$  over a period of  $T$  equally spaced time points. Let  $\mathbf{Z}_t = (Z(s_1, t), \dots, Z(s_n, t))'$  denote the  $n$ -dimensional observation vector at time point  $t, t = 1, \dots, T$ . Then given the true process of interest, the data model takes the form:

$$\mathbf{Z}_t = K\mathbf{O}_t + \varepsilon_{zt} \tag{2.1}$$

where  $\mathbf{O}_t = (O(s_1, t), \dots, O(s_N, t))'$  is an unobserved true process whose components represent the average pollutant concentration for each particular grid cell.  $N$  is the number of grid cells in the domain and  $K$  is a  $n \times N$  mapping matrix that maps averaged pollutant concentration throughout the domain to the observed concentration levels. Thus, the  $K$  matrix elements are zeros and ones and a non-zero element maps the  $i$ th monitoring site to the  $j$ th grid cell (Wikle, 1998). The  $\boldsymbol{\varepsilon}_t$  is a white noise process that represents sub-grid-scale variations and measurement error (nugget effect). Thus we assume that the components of  $\boldsymbol{\varepsilon}_t$  are independent and identically distributed normal random variables with zero mean and unknown variance, that is,  $\boldsymbol{\varepsilon}_t \sim N_n(0, \sigma_z^2 I)$ . Additionally, the model assumes that the mean value of all observations within a grid cell is equal to that grid cell average concentration level represented by the true process. The above assumptions also signify that the spatiotemporal interactions in the data are presented by the true process,  $\mathbf{O}_t$ .

### 2.3.2 Process model

A natural approach for modeling  $O(s, t)$  is by using scientific arguments. Using physical reasoning the pollutant concentration field is further decomposed into three components namely: large-scale spatiotemporal trends, local production and transport (advection). That is:

$$O(s, t) = \mu(s, t) + O^l(s, t) + O^a(s, t) + \eta(s, t) \quad (2.2)$$

where  $\eta(s, t)$  represent the unexplained variability at this stage of the hierarchical model.

In addition, it is known that meteorological variables are significant predictors of a pollutant concentration (Burrows et. al, 1994; McKendry, 2002). Thus, grouping meteorology in two categories of variables related to the above decomposition will enable us

to quantify the significance of these unobservable processes. The first category ( $M^l$ ) relevant to the pollutant local production may include for example: surface temperature, solar radiation, and humidity. The second category ( $M^a$ ) relevant to the transport may include: wind speed and wind direction. Other important meteorological predictors are those related to air mass temperature and convective stability (Burrows et. al, 1995). However, these data are not easily available and for example require coupling the pollution model to a weather prediction system.

The proposed decomposition of the pollutant field identifies the mechanisms (sources of heterogeneity) that contribute to a non-stationary covariance structure. This approach allows development of physically motivated submodels for each of the components conditioned on different categories of meteorological variables.

### 2.3.2.1 Large-scale spatiotemporal trends

I propose a deterministic additive model (multiplicative on the original scale if  $O(s, t)$  are on the logarithmic scale) to capture the data mean structure defined by the equation

$$\mu(s, t) = \alpha(t) + \beta(s) \quad (2.3)$$

in vector form

$$\boldsymbol{\mu} = \boldsymbol{\alpha} \otimes \mathbf{1}_{N \times 1} + \mathbf{1}_{T \times 1} \otimes \boldsymbol{\beta} \quad (2.4)$$

The additive form of the model is physically justifiable since the physical phenomena behind the large-scale temporal and spatial trends are independent. The large scale temporal variability,  $\alpha(t)$ , mostly dominated by the seasonal and diurnal cycles is caused by the planetary motion. On the other hand the large-scale spatial pattern,  $\beta(s)$ , is mostly dominated by the emissions rates of the pollutants or their precursors throughout the do-

main of interest. Heterogeneous emissions rates are one of the mechanisms that induce heterogeneity in the pollutant process and can explain different capacities of sub regions to produce pollutant concentration all other conditions being unchanged. We assume that the emissions are relatively constant in time (i.e.  $\beta(s, t) = \beta(s)$ ), which is in contrast to the other two types of heterogeneity modeled by transport and local production components of the model.

How we model the large-scale temporal trend,  $\alpha(t)$ , depends on what kind of data we have at hand. There are several approaches to model  $\alpha(t)$  including: Box-Jenkins models (Box and Jenkins, 1976); linear combination of different basic functions such as: cubic spline (Huang and Hsu, 2004); sine and cosine components (Huerta et al. 2004); indicator functions. However, using the difference operator in a Box-Jenkins model for removing seasonality and large scale trends can be dangerous. For complex signals where different scales and types/interactions of signals are presented, differencing could deform the remaining signal in unpredictable ways and lead to wrong interpretation and conclusions. For example, to remove the linear trend in a non stationary random time series  $X_t = at + b + \varepsilon_t$  one may decide to use first differencing, since

$$Z_t = X_t - X_{t-1} = at + b - [a(t-1) + b] + \varepsilon_t = a + \varepsilon_t,$$

but if the signal is more complex such as  $X_t = at + b + \sin(\pi t/k) + \varepsilon_t$ . The first differencing leads to  $Z_t = 2 \cos\left(\frac{\pi(2t-1)}{2k}\right) \sin\left(\frac{\pi}{2k}\right) + a + \varepsilon_t$ , a residual signal that has different phase, amplitude (smaller as k is large) and period than the one of the original signal.

The hourly mean data are the most difficult to model since they contain several deterministic signals such as: annual trends, diurnal and seasonal cycles and unknown short term fluctuations. In Chapter 4 I give an example of how the large-scale temporal variability,  $\alpha(t)$ , for the Metro Vancouver dataset can be modeled.

The large-scale spatial pattern due to varying emission rates can have a very complex structure. That structure sometimes cannot be captured easily by traditional methods

such as polynomial function of latitude and longitude or Markov Random Fields (MRF) (Wikle et al. 1998; Royle and Berliner, 1999). Alternatively, the emissions structure can be estimated from the data by the first spatial pattern of a PCA. Calculating the spatial patterns directly from observations will include heterogeneity induced by the transport (wind). This problem can be relatively easily eliminated by selecting only observations that correspond to very light winds (e.g.  $< 3 \text{ km/h}$ ) into the analysis.

The important point here is that we take the mean field as a measure of the emissions rate structure and we use the first  $EOF_1$  mode in our analysis since this eigenfunction resembles the mean surface. The mean field is obtained by averaging out the different meteorological conditions on which the pollutant concentration levels depend. As a result averaging eliminates non-stationary mechanisms which are responsible for local production and what remains is a spatial pattern caused by the varying emissions rates. One can use the mean field to estimate the large-scale emissions structure as well but we think that it is safer to use  $EOF_1$  which allow us to not ignore some other potential sources of non-stationarity. We should also notice that the emission rates alternatively can be obtained from emissions inventory data.

The view outlined above leads us to the following model for the mean spatial structure

$$\beta(s) = b \text{ EOF}_1(s) \tag{2.5}$$

where  $b$  is the proportionality parameter that should be estimated from the data.

### 2.3.2.2 Transport model

Again, we can use physical arguments to build a statistical model to evaluate the regional transport of an air pollutant. Specifically, since the mean wind flow is responsible for the redistribution of the pollutant by advection, we can use the advection equation, which

relates the local changes of a pollutant concentrations levels to the mean wind flow. That is:

$$\frac{\partial O(x, y; t)}{\partial t} = U(x, y; t) \frac{\partial O(x, y; t)}{\partial x} + V(x, y; t) \frac{\partial O(x, y; t)}{\partial y} \quad (2.6)$$

where  $(u, v)$  represent the east-west and north-south components respectively of the mean wind flow and  $(x, y)$  are the corresponding directions.

Following an approach originating in the study of dynamic systems modeling with random perturbations (see for example Pontriagin et al. (Понтрягин и др.), 1933; Christakos, 1991) we can easily approximate (2.6) using finite difference methods. For example, we can construct an explicit scheme for this equation; using first central difference approximation for the first space derivative and first forward difference for the time derivative (Kalnay, 2003) and we get

$$\begin{aligned} O((k, l), t) - O((k, l), t - 1) \approx \\ u((k, l), t - 1)[O((k + 1, l), t - 1) - O((k - 1, l), t - 1)] \\ + v((k, l), t - 1)[O((k, l + 1), t - 1) - O((k, l - 1), t - 1)] \end{aligned} \quad (2.7)$$

Alternatively, if we use a fully implicit (backward) time scheme we get

$$\begin{aligned} O((k, l), t) - O((k, l), t - 1) \approx u((k, l), t)[O((k + 1, l), t) - O((k - 1, l), t)] \\ + v((k, l), t)[O((k, l + 1), t) - O((k, l - 1), t)] \end{aligned} \quad (2.8)$$

where  $(k, l)$  represent the grid cell coordinates. The first approximation (2.7) suggests lagged spatial-time interaction on the other hand, the second one approximation (2.8) suggests simultaneous spatial-time interaction. Similarly, both (2.7) and (2.8) suggest an autoregressive term. In the context of the advection equation (2.6) the autocorrelation can



be seen as non-wind transport (persistence) from the previous day. The time autocorrelation which naturally arises from (2.6) has been employed in modeling of numerous air pollution problems: Sahu et al., 2007; McMillan et al., 2005 etc.

The explicit scheme (2.7) suggests the following stochastic transport model that can be used to explain much of the spatiotemporal structure of the  $O(s, t)$  process

$$\begin{aligned}
O^a(k, l, t) = & \rho O(k, l, t - 1) \\
& + \theta_1 u(k, l, t - 1) [O(k + 1, l, t - 1) - O(k - 1, l, t - 1)] \\
& + \theta_2 v(k, l, t - 1) [O(k, l + 1, t - 1) - O(k, l - 1, t - 1)], \tag{2.9}
\end{aligned}$$

$t = 2, \dots, T$

where spatial autoregressive parameters  $\theta = (\theta_1, \theta_2)$  depend on the lattice specification, have no dimensions, and are characterized by the spatial pollutant dispersion in east-west and north-south directions of the region. For a small flat region it is reasonable to assume that  $\theta$ 's have no spatial structure.  $\rho$  - denotes the time autoregressive parameter.

The vector form of the model (9) is given by

$$\mathbf{O}_t^a = \rho \mathbf{O}_{t-1} + H(\mathbf{U}_{t-1}, \mathbf{V}_{t-1}; \boldsymbol{\theta}) \mathbf{O}_{t-1} \tag{2.10}$$

where  $H$  is a four, off diagonal  $N \times N$  matrix of regression coefficients, parameterized to allow the first-order lagged nearest-neighbor structure described above,  $\boldsymbol{\theta}$  is the parameters vector, and  $(\mathbf{U}_{t-1}, \mathbf{V}_{t-1})$  are  $N$ -vectors of the  $u$  and  $v$  components of the wind field at  $N$  grid locations.

The fully implicit scheme (2.8) suggests the following stochastic transport model

$$O^a(k, l, t) = \rho O(k, l, t - 1) \tag{2.11}$$

$$\begin{aligned}
& +\theta_1 u(k, l, t)[O((k+1, l), t) - O((k-1, l), t)] \\
& +\theta_2 v(k, l, t)[O((k, l+1), t) - O((k, l), t)], \quad t = 2, \dots, T.
\end{aligned}$$

The vector form of model (11) is given by

$$\mathbf{O}_t^a = \rho \mathbf{O}_{t-1} + H(\mathbf{U}_t, \mathbf{V}_t; \boldsymbol{\theta}) \mathbf{O}_t \quad (2.12)$$

or using (2.2), we obtain

$$(I - H)(\mathbf{O}_t - \boldsymbol{\mu}_t - \rho \mathbf{O}_{t-1} - \mathbf{O}_t^l) = \boldsymbol{\eta}_t \quad (2.13)$$

where  $\mathbf{O}_t^l$  and  $\boldsymbol{\mu}_t$  are  $N$ -vectors of the local production component and the mean process at  $N$  grid locations. Again,  $H$  is a four, off diagonal  $N \times N$  matrix of regression coefficients, but this time parameterized to allow the first-order simultaneous nearest-neighbor structure described above.

Which model specification is preferable depends on the data at hand. For hourly data the lagged model (2.10) is more appropriate. However, for daily maximum 8-hour average data the simultaneous model (2.12) is preferable. Given that we have the wind field, it appears to be more realistic to associate today's transport effect at given location with today's values at its neighboring locations than yesterday's values. A possible justification for such an assumption would be that the time scale of daily maximum 8-hour ozone levels is hours (as is the ground-level horizontal ozone transport effect) but not days.

### 2.3.2.3 Local production model

In this section we will consider a local production model for pollutants whose generating mechanism depends mostly on the meteorology. We would not need such a model if we were dealing with a primary air pollutant. In this case our focus will be the emissions model. However, if we model a secondary air pollutant such as ozone we will need to develop a model conditional on the local meteorology.

What are reasonable requirements the local production model should meet? At the first place the model should ignore dynamic contributions from the neighboring sites and take into account only specific local meteorological conditions, that is, we should allow covariates to be functions of location, and regression coefficients should not depend on location. On one hand the spatially dependent regression coefficients will account for part of the spatial heterogeneity induced by the transport processes. In this way, our local production model will explain part of the pollutant transport and will therefore not be a fully local term. We want this kind of heterogeneity to be modeled only by the transport model and only there. On the other hand this will put physical/chemical meaning in our regression coefficients, which is a desirable feature of a local production model since the physics and chemistry do not depend on the location. Furthermore, if we decide to use regionally averaged covariate fields it will mean that we treat the response process (pollutant concentration) differently from its covariates (different meteorological fields). This is a scale problem since the averaged covariates are surrogates at a coarser scale of the responses that occur at a finer scale. An important feature of this approach is that the local production spatial structure is not modeled directly, but inherent in the covariates field. This requirement is also reasonable since local production is strongly dependent on the meteorological fields.

Taking into account the above reasoning, we propose the following regression model for modeling the local production:

$$O^l(s_i, t) = M(s_i, t)\beta \quad (2.14)$$

or in vector form

$$\mathbf{O}_t^l = M_t \boldsymbol{\beta} \quad (2.15)$$

where  $M_t$  is  $N \times k$  design matrix of the exploratory variables at time point  $t$ , and  $\boldsymbol{\beta}$  is a  $k$ -vector of random regression parameters. Physically, these parameters signify the relative importance of the different meteorological fields that are relevant to local pollutant production. In particular, in Chapter 4 we will use the following meteorological fields to estimate ozone local production: temperature, relative humidity and solar radiation. For simplicity the proposed model is linear even if the local production processes are not linear. However, it has been shown in several studies (Gardner and Dorling, 2000) that linear regression models produce very similar results to those of non linear models.

Substituting, the latent processes  $\mathbf{O}^a$  and  $\mathbf{O}^l$  in (2.2) by (2.10) and (2.15) we obtain the following process model:

$$\begin{aligned} \mathbf{O}_t &= \boldsymbol{\mu}_t + M_t \boldsymbol{\beta} + \rho \mathbf{O}_{t-1} + H(\mathbf{U}_{t-1}, \mathbf{V}_{t-1}; \boldsymbol{\theta}) \mathbf{O}_{t-1} + \boldsymbol{\eta}_t, \\ &t = 2, \dots, T \end{aligned} \quad (2.16)$$

where  $\{\boldsymbol{\eta}_t\}$  is a collection of independent unstructured error terms specifically assumed to follow:

$$\boldsymbol{\eta}_t \sim N_N(0, \sigma_o^2 I) \quad (2.17)$$

The autoregressive model (2.16) requires an initial condition for  $\mathbf{O}_1$  to be specified. We assume the following model:

$$\mathbf{O}_1 = \boldsymbol{\mu}_1 + M_1\boldsymbol{\beta} + \boldsymbol{\eta}_1 \quad (2.18)$$

We should also take care in dealing with points on the boundary of the region. In such cases there will not be observations on all sides of a given point. In Chapter 3 we will consider two examples about how we can deal with this problem.

Finally, because identifiability of the variance components of  $\boldsymbol{\eta}_t$  in (2.2) the covariance structure of the proposed submodels in this section is free of a nugget effect, that is, we do not introduce error terms in the equations (2.10) and (2.15), otherwise it is assumed that the overall nugget effect can be decomposed in the same manner as the true process,  $\mathbf{O}_t$ , in known proportions.

### 2.3.3 Parameters model

To complete the hierarchical model we need to specify the prior distribution of the model parameters such as: variance parameters describing measurement error and uncertainty in each level of the hierarchy, meteorological regression parameters, and spatial autoregressive parameters, that is,  $[\boldsymbol{\theta}, \boldsymbol{\beta}, \rho, \sigma_z^2, \sigma_o^2]$ . In general, we will assume an independent parameters model i.e.

$$[\boldsymbol{\theta}, \boldsymbol{\beta}, \rho, \sigma_z^2, \sigma_o^2] = [\boldsymbol{\theta}][\boldsymbol{\beta}][\rho][\sigma_z^2][\sigma_o^2]. \quad (2.19)$$

the selected priors should lead to meaningful proper posteriors and allow us to incorporate relevant physics into the model. We specify the following conjugate priors:

$$\begin{aligned} [\boldsymbol{\theta}] &\sim N_2(\tilde{\boldsymbol{\theta}}, \tilde{\sigma}_{\boldsymbol{\theta}}^2 I), & [\boldsymbol{\beta}] &\sim N_3(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}_{\boldsymbol{\beta}}^2 I), & [\rho] &\sim N(\tilde{\rho}, \tilde{\sigma}_{\rho}^2) \\ [\sigma_z^2] &\sim IG(\tilde{\alpha}_z, \tilde{\beta}_z), & [\sigma_o^2] &\sim IG(\tilde{\alpha}_o, \tilde{\beta}_o). \end{aligned} \quad (2.20)$$

The final stage of the hierarchy requires numerical values to be specified for all unknowns  $\tilde{\boldsymbol{\theta}}, \tilde{\sigma}_{\theta}^2, \tilde{\boldsymbol{\beta}}, \tilde{\sigma}_{\beta}^2, \tilde{\rho}, \tilde{\sigma}_{\rho}^2, \tilde{\alpha}_z, \tilde{\beta}_z, \tilde{\alpha}_o, \tilde{\beta}_o$ .

## 2.4 Prediction

Let  $O = (\mathbf{O}_1, \dots, \mathbf{O}_T)$ , so that, the  $O$  represents the unobserved true process of interest over the study period. We define  $Z$  and  $M$  in a similar fashion, where  $Z$  denotes observed pollutant levels and  $M$  denotes the observed meteorological fields. The joint posterior distribution of the all unknowns in the hierarchical model specified by (2.1), (2.16) and (2.19) is given by:

$$[\boldsymbol{\theta}, \boldsymbol{\beta}, \rho, O, \sigma_z^2, \sigma_o^2, Z, M] \propto [\boldsymbol{\theta}][\boldsymbol{\beta}][\rho][\sigma_z^2][\sigma_o^2][\mathbf{O}_1|M_1, \mu_1, \beta, \sigma_o^2][Z_1|K, O_1, \sigma_z^2] \prod_{t=2}^T [\mathbf{O}_t|\mathbf{O}_{t-1}, \boldsymbol{\theta}, \boldsymbol{\beta}, \rho, \mu, \sigma_o^2, H_t, M_t][Z_t|K, \mathbf{O}_t, \sigma_z^2]. \quad (2.21)$$

The full conditional distribution of the unknown variates can now be obtained by collecting the terms of (2.20) that correspond to the variate of interest (Gilks et al., 1996).

We also use the joint posterior distribution (2.21) to make k-step ahead predictions. Let  $\xi$  denotes the collection of all unknowns, such that  $\xi = \{\boldsymbol{\theta}, \boldsymbol{\beta}, \rho, O, \sigma_z^2, \sigma_o^2\}$ . The one-step-ahead prediction distribution is given by

$$[\mathbf{O}_{T+1}|O, Z, M] = \int [\mathbf{O}_{T+1}|\xi][\xi|O, Z, M]d\xi \quad (2.22)$$

where the likelihood term  $[O_{T+1}|\xi]$  is given by (2.16) assuming that the meteorology at the time moment  $T + 1$  is given. Using the squared error loss the optimal one-step-ahead forecast is given by  $E[O_{T+1}|Z, M]$ . We approximate  $E[O_{T+1}|Z, M]$  by drawing samples  $\mathbf{o}_{T+1}^{(j)}$  from  $[\mathbf{o}_{T+1}|\xi^{(j)}]$  and forming the sample average (Banerjee, 2004).

## CHAPTER 3

### SIMULATION STUDY

In this chapter, we examine the performance of the model presented in Chapter 2 using different scenarios while testing the model code implementation. Of particular interest is how the parameter inference and spatial interpolation depend on: the way we complete the gradients in (2.6) on the boundary of the region; whether we allow the time autoregressive process to vary spatially; available data; prior information.

#### 3.1 Boundary sites

Let us consider the following first-order neighbor structure which arises from the gradients approximation in (2.7) or (2.8)

$$\begin{array}{ccccc} & & s_2 & & \\ & & | & & \\ s_1 & & s_3 & & s_5 \\ & & | & & \\ & & s_4 & & \end{array} .$$

Apparently, if the point  $s_3$  is on the boundary of the domain of interest one or two of its neighbors will be missing. This requires special treatment of the boundary sites. In this simulation study we consider two different methods for dealing with this problem.

In the first approach, we use only available observations and complete the gradients by using only the data from the central point ( $s_3$ ). For example, if  $s_3$  is located at the left bottom corner of the domain there will be two missing neighbor sites one to the west and one to the south. In this case the East-West gradient will given approximate by



$O(s_5, \cdot) - O(s_3, \cdot)$  and the North-South gradient will given approximate by  $O(s_2, \cdot) - O(s_3, \cdot)$ , Royle et al., 1998.

In the second approach, following McMillan (2005) we can extend the domain by introducing boundary sites and their associated processes. That is, we consider a collection of four boundary processes  $\{\mathbf{O}_{it}^B\}$  for each time point  $t$  and for each side  $i = (1, \dots, 4)$  of the domain. The  $\mathbf{O}_{it}^B$  are  $N \times 1$  vectors whose elements correspond to the sites needed to complete the boundary gradients. We model the  $\mathbf{O}_{it}^B$  stochastically which allows us to account for uncertainty and dependence between the neighboring boundary sites; see Wikle et al. (2003) for general discussion.

In this case, the process model given by (2.16) should be extended by adding the following term

$$\mathbf{o}_t^B = \sum_{i=1}^4 H_i^B \mathbf{o}_{it}^B \quad (3.1)$$

and process model will take the form

$$\mathbf{o}_t = \boldsymbol{\mu}_t + M_t \boldsymbol{\beta} + \rho \mathbf{o}_{t-1} + H(\mathbf{U}_{t-1}, \mathbf{V}_{t-1}; \boldsymbol{\theta}) \mathbf{o}_{t-1} + \sum_{i=1}^4 H_i^B \mathbf{o}_{it}^B + \boldsymbol{\eta}_t, \quad (3.2)$$

$$t = 2, \dots, T$$

and

$$\mathbf{o}_1 = \boldsymbol{\mu}_1 + M_1 \boldsymbol{\beta} + \sum_{i=1}^4 H_i^B \mathbf{o}_{i1}^B + \boldsymbol{\eta}_1, \quad (3.3)$$

for  $t = 1$ . The  $H_i^B$  are  $N \times N$  matrices which maps the  $\mathbf{O}_{it}^B$  to the corresponding sites with incompleted gradients. For example, if we count the grid cells from the bottom left corner of the grid. The north boundary matrix  $H_1^B$  is given by

$$H_1^B = \begin{pmatrix} 0 & & \dots & 0 & \dots & 0 \\ \vdots & & & & & \\ 1 & & & \vdots & & \vdots \\ & \ddots & & & & \\ 0 & & 1 & 0 & \dots & 0 \end{pmatrix}$$

and the west boundary matrix  $H_4^B$  is given by

$$H_4^B = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & & \vdots & & & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix}.$$

We model the boundary processes  $\mathbf{O}_{it}^B$  hierarchically as well

$$\begin{aligned} \mathbf{O}_{it}^B &= \mathbf{v}_{it} + \boldsymbol{\varepsilon}_{it}^B, \quad i = 1, \dots, 4 \\ \boldsymbol{\varepsilon}_{it}^B &\sim N_N(0, \sigma_B^2 I) \\ \mathbf{v}_{it} &= \phi_{i1} \mathbf{1} + \phi_{i2} \mathbf{v}_{it-1} + \boldsymbol{\varepsilon}_{it}^v, \quad t = 2, \dots, T \\ \boldsymbol{\varepsilon}_{it}^v &\sim N(0, \sigma_{v_i}^2 I) \\ \mathbf{v}_{i1} &= \phi_{i0} \mathbf{1} + \boldsymbol{\varepsilon}_{i1}^{v_1}, \\ \boldsymbol{\varepsilon}_{i1}^{v_1} &\sim N(0, \sigma_{v_0}^2 I). \end{aligned} \tag{3.4}$$

The final stage of the model requires specification of the prior distributions of the variance parameters  $\sigma_B^2, \sigma_{v_i}^2$ , and  $\sigma_{v_0}^2$  as well as of the autoregressive parameters  $\phi_{i1}, \phi_{i2}$ , and  $\phi_{i0}$ . Again, we assume that all of these parameters are independent. For the variance parameters we use the conjugate inverse gamma prior, that is

$$\begin{aligned}\sigma_B^2 &\sim IG(\tilde{\alpha}, \tilde{\beta}) \\ \sigma_{v_i}^2 &\sim IG(\tilde{\alpha}_{v_i}, \tilde{\beta}_{v_i}), \quad i = 1, \dots, 4 \\ \sigma_{v_{0i}}^2 &\sim IG(\tilde{\alpha}_{v_{0i}}, \tilde{\beta}_{v_{0i}}).\end{aligned}\tag{3.5}$$

To simplify matters we assume the same variances for the boundary rows i.e.  $\sigma_r^2 = \sigma_{v_1}^2 = \sigma_{v_3}^2$ ,  $\sigma_{r0}^2 = \sigma_{v_{01}}^2 = \sigma_{v_{03}}^2$ , and for the boundary columns i.e.  $\sigma_c^2 = \sigma_{v_2}^2 = \sigma_{v_4}^2$ ,  $\sigma_{c0}^2 = \sigma_{v_{02}}^2 = \sigma_{v_{04}}^2$ . For the autoregressive parameters we use the conjugate normal prior, that is

$$\phi_{ij} \sim N(\tilde{\phi}_{ij}, \tilde{\sigma}_{ij}^2), \quad i = 1, \dots, 4; \quad j = 0, \quad 1, \quad 2.\tag{3.6}$$

and simplifying assumptions  $\phi_{rj} = \phi_{1j} = \phi_{3j}$  and  $\phi_{cj} = \phi_{2j} = \phi_{4j}$  and  $\tilde{\sigma}_\phi = \tilde{\sigma}_{ij}$ .

### 3.2 The autoregressive process

In this chapter we also investigate how the model performance depends on the autoregressive process  $\rho$ . Initially we do not allow this process to vary spatially or temporally, that is,

$$\rho(s, t) = \rho \text{ and } \rho \sim N(\mu_\rho, \sigma_\rho^2).\tag{3.7}$$

Next, we will allow spatially varying autoregressive processes but do not allow spatial dependence among their components i.e. we will allow only independent local adjustments. Formally, this can be written as

$$\boldsymbol{\rho} \sim N(\boldsymbol{\mu}_\rho, \sigma_\rho^2 \mathbf{I}). \quad (3.8)$$

These assumptions seem to be realistic since we expect the large-scale spatio-temporal effects to be captured by the deterministic mean process,  $\boldsymbol{\mu}$ , and eventually we need only independent local adjustments to capture within grid-scale variability.

Alternatively, following Wikle et al. (1998) we can model the spatially varying autoregressive parameters  $\boldsymbol{\rho}$ , as a conditionally specified Gaussian MRF with first-order spatial dependence

$$\begin{aligned} [\rho(k, l) | \rho(i, j): (i, j) \neq (k, l)] \sim & N(\rho_0(k, l) + \alpha_\rho\{\rho(k-1, l) - \rho_0(k-1, l) \\ & + (\rho(k+1, l) - \rho_0(k+1, l))\} + \beta_\rho\{\rho(k, l-1) \\ & - \rho_0(k, l-1) + (\rho(k, l+1) - \rho_0(k, l+1))\}, \tau^2) \end{aligned} \quad (3.9)$$

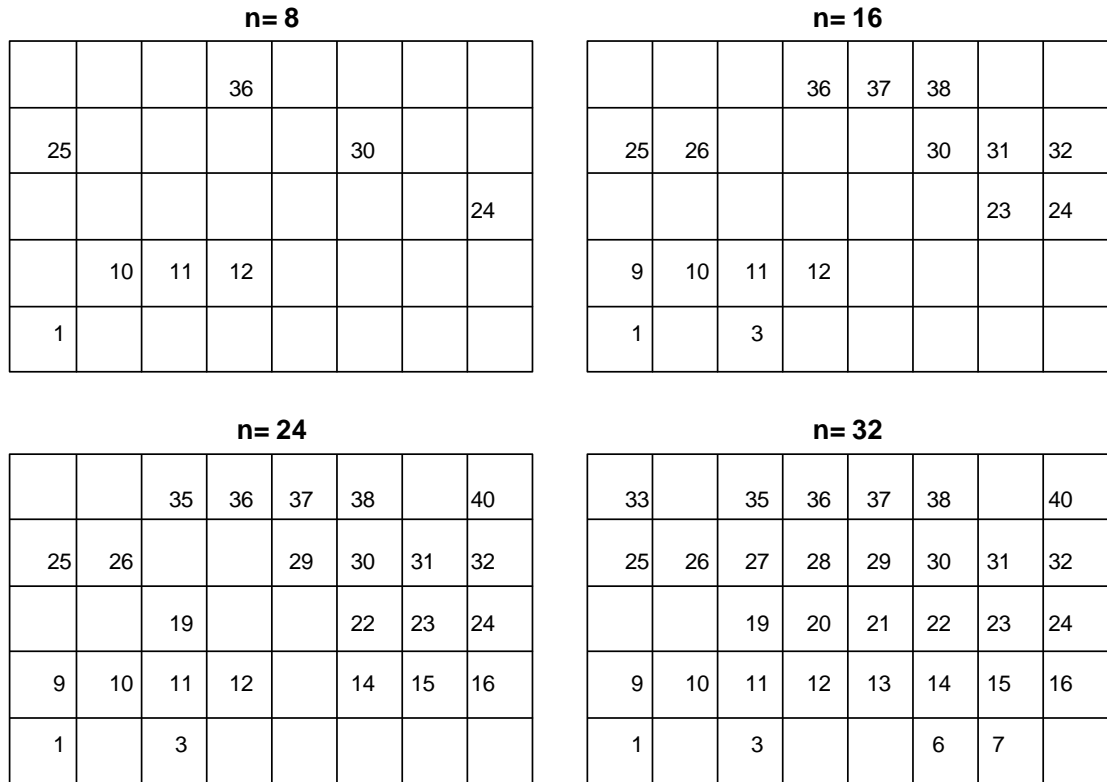
where  $\rho_0$  is the MRF mean,  $\alpha_\rho$  and  $\beta_\rho$  are MRF spatial dependence parameters, and  $\tau^2$  is the homogeneous variance.

### 3.3 Spatial data coverage

For our simulations, we consider a regular lattice of  $G = 5 \times 8$  (so  $N = 40$ ) and  $T = 30$  equally spaced time points. We simulate a realization of the process given by

$$\begin{aligned} \mathbf{Z}_t &= M_t \boldsymbol{\beta} + \rho \mathbf{Z}_{t-1} + H(\mathbf{U}_{t-1}, \mathbf{V}_{t-1}; \boldsymbol{\theta}) \mathbf{Z}_{t-1} + \boldsymbol{\epsilon}_t \\ \boldsymbol{\epsilon}_t &\sim N_N(0, \sigma_Z^2 \mathbf{I}), \quad t = 2, \dots, T \\ \mathbf{Z}_1 &= M_1 \boldsymbol{\beta} + \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_1 &\sim N_N(0, \sigma_Z^2 \mathbf{I}). \end{aligned} \quad (3.10)$$

The difference between the process model (2.16) and (3.10) is that (3.10) does not include large-scale spatiotemporal trends,  $\boldsymbol{\mu}(s, t)$ . This difference is not essential since the statistical inferences are valid for both processes. We compile four different datasets based on the above realization and fit the hierarchical model (2.1), (2.16), and (2.20) given in the previous chapter. The number of grid cells where data are available  $n$  (this is termed spatial data availability) varies from 20% to 80% corresponding to sequence of 8, 16, 24, 32 grid cells. The spatial sampling locations are selected at random and are shown in Fig. 3.1. The meteorological fields  $M_t, U_t, V_t$  were generated from Gaussian processes centered at the observed mean values of the dataset which will be described in Chapter 4 and using corresponding sample variances.



**Figure 3.1** Selected grid cells for the four datasets. Cells are counted from the left bottom corner.

### 3.4 Prior information

To investigate the effect of having different degrees of prior knowledge, we consider two prior specifications for the parameters. For that purpose we specify the hyperparameter values in the prior distributions, (2.20), (3.5), and (3.6) to convey different degrees of information.

In our first specification given in Table 3.1 we consider proper noninformative priors. Noninformative prior distributions are intended to allow Bayesian inference for parameters about which we have no previous information.

**Table 3.1** Summary of the noninformative prior specification.

Parameter	Prior Mean	Prior Variance	Prior Parameters
$\sigma_z^2$	30	100	$\tilde{\alpha}_z = 11, \tilde{\beta}_z = 300$
$\sigma_o^2$	30	100	$\tilde{\alpha}_o = 11, \tilde{\beta}_o = 300$
$\sigma_B^2$	1	10	$\tilde{\alpha}_B = 2.1, \tilde{\beta}_B = 1.1$
$\sigma_{v_1}^2, \sigma_{v_2}^2, \sigma_{v_{01}}^2, \sigma_{v_{02}}^2$	0.1	10	$\tilde{\alpha}_v = 2.001, \tilde{\beta}_v = 0.1$
$\boldsymbol{\beta}$	(0,0,0)	$diag(\tilde{\sigma}_\beta^2 \mathbf{1}_3)$	$\tilde{\boldsymbol{\mu}}_\beta = 0, \tilde{\sigma}_\beta^2 = 10000$
$\boldsymbol{\theta}$	(0,0)	$diag(\tilde{\sigma}_\theta^2 \mathbf{1}_2)$	$\tilde{\boldsymbol{\mu}}_\theta = 0, \tilde{\sigma}_\theta^2 = 10000$
$\boldsymbol{\rho}$	$\mathbf{0}_N$	$diag(\tilde{\sigma}_\rho^2 \mathbf{1}_N)$	$\tilde{\boldsymbol{\mu}}_\rho = 0, \tilde{\sigma}_\rho^2 = 10000$
$\phi_{ij}$	0	1000	$\tilde{\mu}_\phi = 0, \tilde{\sigma}_\phi^2 = 1000$

In the second specification, the informative prior, we use the same priors for the boundary related parameters, that is,  $\sigma_{v_1}^2, \sigma_{v_2}^2, \sigma_{v_{01}}^2, \sigma_{v_{02}}^2, \phi_{ij}$ , and  $\sigma_B^2$ , but priors for

$\sigma_z^2, \sigma_o^2, \boldsymbol{\theta}, \boldsymbol{\beta}, \rho$  are centered very close to the true values. Specifically, Table 3.2 gives the specification we use for non-spatially varying autoregressive process,  $\rho$ .

**Table 3.2** Summary of the informative prior specification.

Parameter	Prior Mean	Prior Variance	Prior Parameters
$\sigma_z^2$	1	10	$\tilde{\alpha}_z = 2.1, \tilde{\beta}_z = 1.1$
$\sigma_o^2$	1	10	$\tilde{\alpha}_o = 2.1, \tilde{\beta}_o = 1.1$
$\sigma_B^2$	1	10	$\tilde{\alpha}_B = 2.1, \tilde{\beta}_B = 1.1$
$\sigma_{v_1}^2, \sigma_{v_2}^2, \sigma_{v_{01}}^2, \sigma_{v_{02}}^2$	0.1	10	$\tilde{\alpha}_v = 2.001, \tilde{\beta}_v = 0.1$
$\boldsymbol{\beta}$	(0.4, 0.2, 0.01)	$diag(\tilde{\sigma}_\beta^2 \mathbf{1}_3)$	$\tilde{\sigma}_\beta^2 = 10$
$\boldsymbol{\theta}$	(0.1, 0.15)	$diag(\tilde{\sigma}_\theta^2 \mathbf{1}_2)$	$\tilde{\sigma}_\theta^2 = 10$
$\rho$	0.4	$diag(\tilde{\sigma}_\rho^2 \mathbf{1}_N)$	$\tilde{\sigma}_\rho^2 = 10$
$\phi_{ij}$	0	1000	$\tilde{\mu}_\phi = 0, \tilde{\sigma}_\phi^2 = 1000$

### 3.5 Results

We consider four different forms of the hierarchical model described in the previous chapter to fit the simulated realization of the process defined by equations (3.10).

- i) *Model<sub>1</sub>*: the time autoregressive parameter  $\rho$  is not allowed to vary spatially and we use only available observations to evaluate the gradients on the boundaries. That is, we use equation (2.16) as the second stage in our hierarchical model.
- ii) *Model<sub>2</sub>*: the time autoregressive parameter  $\rho$  is allowed to vary spatially and we use only available observations to evaluate the gradients on the boundaries. That is, we use equation (2.16) again as the second stage in our hierarchical model but in this case the time autoregressive parameter  $\rho$  is replace by  $N \times N$  diagonal matrix.

- iii) *Model<sub>3</sub>*: the time autoregressive parameter  $\rho$  is allowed to vary spatially and we extend the domain by introducing boundary sites and their associated processes. That is, we use equation (3.2) and (3.3) as the second stage in our hierarchical model. In this case the autoregressive parameter  $\rho$  is again an  $N \times N$  diagonal matrix. Note that *Model<sub>1</sub>* and *Model<sub>2</sub>* are special cases of model *Model<sub>3</sub>*
- iv) *Model<sub>4</sub>*: the time autoregressive parameter  $\rho$  is not allowed to vary spatially and we extend the domain by introducing boundary sites and their associated processes. That is, we again use equation (3.2) and (3.3) as the second stage in our hierarchical model but in this case, the autoregressive parameter  $\rho$  is a scalar quantity.

All results were obtained via MCMC computer simulations. For simplicity, only the full conditional distributions required for the Gibbs Sampler for the most complicated model, *Model<sub>3</sub>*, are presented in Appendix A.

### 3.5.1 Parameter inference

Tables 3.3, and 3.4 present a summary of the parameters inferred by the model *Model<sub>1</sub>* using two different prior specifications. The results in these tables suggest the following conclusions:

- The regression parameters  $(\beta_1, \beta_2, \beta_3)$  are reasonably well estimated for all four cases, especially  $\beta_3$  which is essentially equal to the true value. The uncertainty about parameters  $(\beta_1, \beta_2)$  declines with the increase of the available observations, and for all datasets the estimates of these parameters are not very sensitive to the prior specification. Figure 3.2 shows MCMC samples from the posterior distributions of these parameter under noninformative prior and  $n = 16$ .



- The mean posterior estimates for the spatial autoregressive parameters  $(\theta_1, \theta_2)$  show poor performance and relatively large standard deviations for a small number of available observations ( $n = 8, 16$ ) in both cases of prior specification. However, the situation substantially improves as the number of the observations increases.
- The time autoregressive parameter  $\rho$  is the most difficult parameter to estimate. The situation does not improve as the number of the available grid cells increases and there is only slight improvement when an informative prior specification is used. This could indicate poor modeling assumptions which we will try to address in our other model versions *Model<sub>2</sub>*, and *Model<sub>3</sub>* that include a spatially varying autoregressive process.
- The inferences about  $\sigma_z^2$  and  $\sigma_o^2$  under the noninformative prior specification are very poor. The data variability is split more or less equally between the first two stages of the hierarchical model indicating poor overall model performance. However, there is significant improvement when we use an informative prior specification.

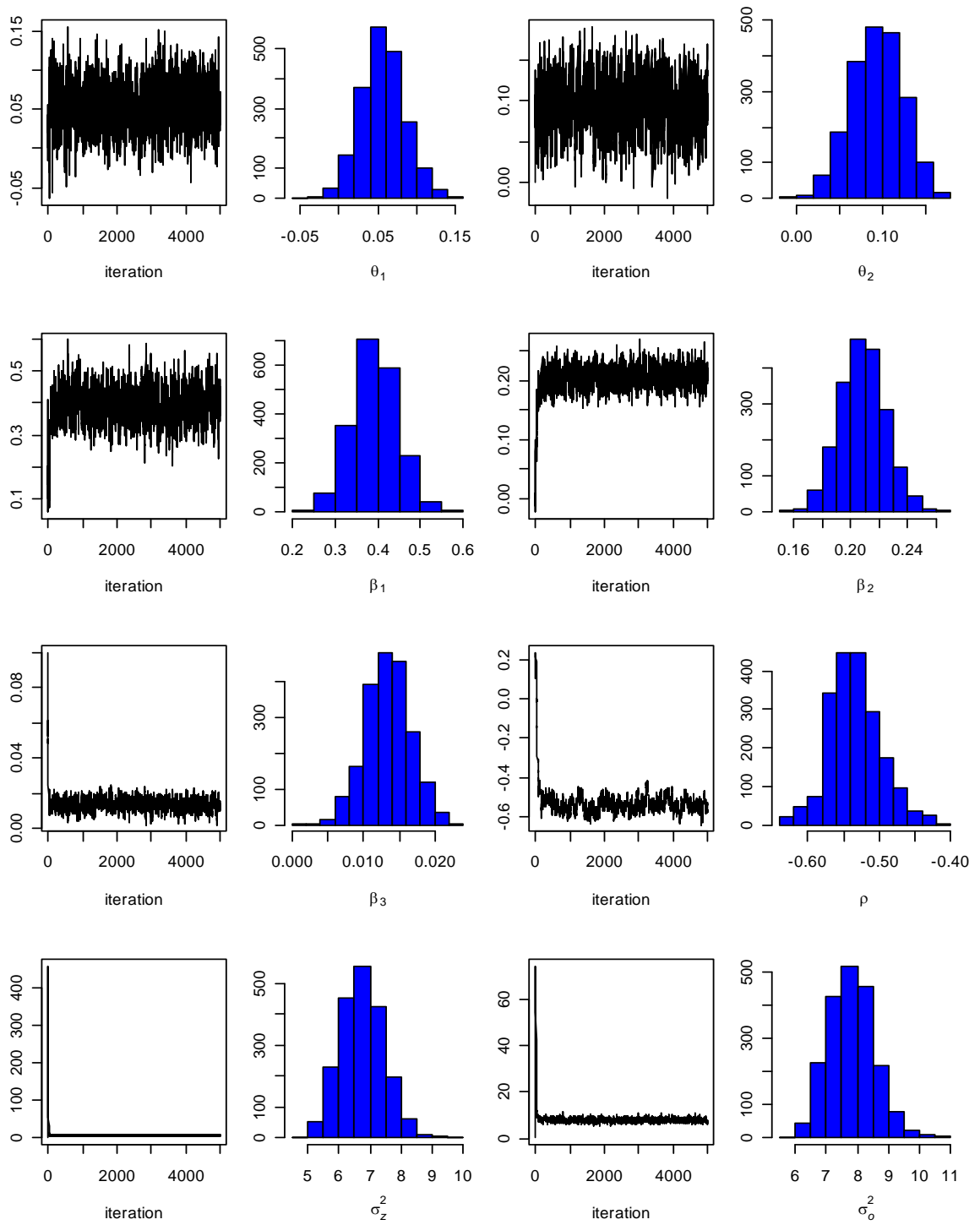
**Table 3.3** Summary of the posterior parameter inference with different numbers  $n$  of available observations based on model *Model<sub>1</sub>* and noninformative prior (Table 3.1).

Parameter	True Value	$n = 8$		$n = 16$		$n = 24$		$n = 32$	
		Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
$\beta_1$	0.4	0.41	0.11	0.39	0.05	0.44	0.04	0.43	0.03
$\beta_2$	0.2	0.2	0.03	0.21	0.02	0.21	0.01	0.21	0.01
$\beta_3$	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.01	0.00
$\theta_1$	0.1	0.02	0.05	0.06	0.03	0.05	0.02	0.08	0.02
$\theta_2$	0.15	0.06	0.05	0.09	0.03	0.13	0.02	0.14	0.02
$\rho$	-0.4	-0.54	0.04	-0.53	0.04	-0.52	0.03	-0.52	0.02
$\sigma_z^2$	10	9.08	1.188	6.76	0.70	5.89	0.55	5.27	0.46
$\sigma_o^2$	NA	10.034	1.282	7.82	0.74	7.42	0.62	7.24	0.54

**Table 3.4** Summary of posterior parameter inference base on model  $Model_1$  using different numbers  $n$  of available observations and informative prior (Table 3.1).

Parameter	True Value	$n = 8$		$n = 16$		$n = 24$		$n = 32$	
		Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
$\beta_1$	0.4	0.39	0.07	0.38	0.04	0.43	0.03	0.41	0.02
$\beta_2$	0.2	0.2	0.02	0.20	0.01	0.20	0.01	0.20	0.01
$\beta_3$	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
$\theta_1$	0.1	0.00	0.04	0.06	0.02	0.06	0.02	0.08	0.02
$\theta_2$	0.15	0.09	0.04	0.12	0.02	0.13	0.02	0.14	0.01
$\rho$	-0.4	-0.51	0.04	-0.49	0.03	-0.47	0.03	-0.48	0.02
$\sigma_z^2$	10	0.86	0.67	1.29	0.80	0.80	0.42	0.7	0.32
$\sigma_o^2$	NA	11.17	1.34	9.45	1.14	10.1	0.77	10.05	0.63

Figure 3.2 displays the MCMC samples from the posterior distributions of all model  $Model_1$  parameters in the case of noninformative prior and number of available grid cells equal  $n = 16$ . We can see that the Markov chains exhibit the same qualitative behavior after a few iterations, which is an indication of convergence. Furthermore, Figure 3.2 also presents the MCMC histograms that approximate the marginal posterior distributions of the model parameters. It is clear from the presented histograms that the model parameters are concentrated around their true values and the algorithm correctly identifies the parameters.



**Figure 3.2** Sequences of 5000 iterations from the posterior distributions of the model  $Model_1$  parameters and the histograms of the last 2000 iterations.

Next, we will concentrate on parametric inference for the model  $Model_2$ .  $Model_2$  has significantly more parameters than models  $Model_1$  and  $Model_4$ , which results in a slower Gibbs sampler and a considerable rise in computation cost. Tables 3.5, and 3.6 present a summary of the inference about the model  $Model_2$  parameters using two different prior specifications. The results in these tables suggest:

- Again, the regression parameters  $(\beta_1, \beta_2, \beta_3)$  are reasonably well estimated for all cases. The uncertainty about these parameters declines with an increase in the available observations, and for all datasets the estimates of these parameters are not very sensitive to the prior specification.
- The spatial autoregressive parameters  $(\theta_1, \theta_2)$  are very sensitive to the number of available observations in both cases of prior specification. The increasing number of observations substantially improves parameter inference.
- The variances  $\sigma_z^2$  and  $\sigma_o^2$  are the parameters that are most strongly influenced by the prior specification and are not very sensitive to the number of available observations.

Figure 3.3 displays the true and interpolated surfaces of the spatially varying autoregressive process,  $\rho$ , for informative prior and  $n = 32$ . A visual inspection of these two surfaces shows that they are closely similar (spatial pattern). To quantify the differences between the true and the model-estimated surfaces of the process  $\rho$  we use a statistic Surface Distance (SD) defined as the mean sum of squared differences between the observed and estimated values

$$SD = \frac{1}{N} \sum_{i=1}^N (\rho_i - \hat{\rho}_i)^2$$

where  $\rho_i$  is the observed and  $\hat{\rho}_i$  estimated values of the  $\rho$  at  $i \in G$ . Table 3.7 reports the SD statistics for  $\rho$ . The SD statistic appears to decrease with increasing available data in a non monotonic fashion.

**Table 3.5** Summary of posterior parameter inference based on model  $Model_2$ , using different number  $n$  of available observations and noninformative prior (Table 3.1).

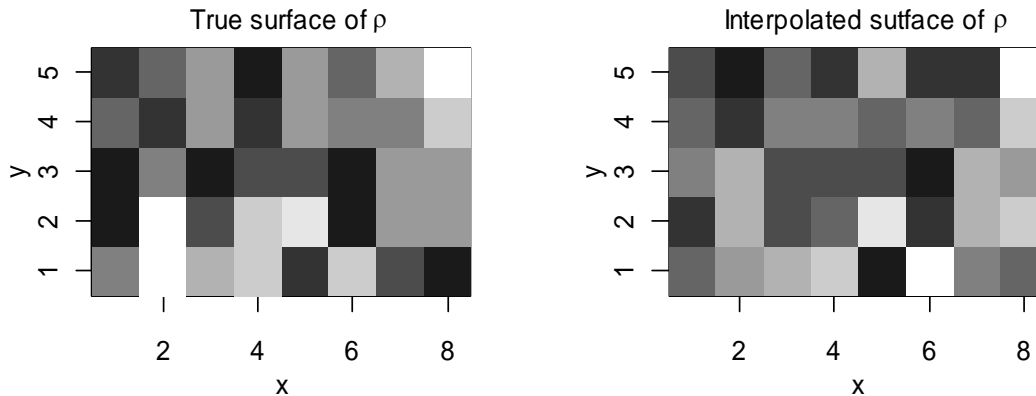
Parameter	True Value	$n = 8$		$n = 16$		$n = 24$		$n = 32$	
		Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
$\beta_1$	0.4	0.5	0.28	0.36	0.13	0.36	0.11	0.42	0.08
$\beta_2$	0.2	0.27	0.1	0.23	0.04	0.23	0.03	0.22	0.03
$\beta_3$	0.01	0.0	0.01	0.01	0.0	0.01	0.0	0.01	0.0
$\theta_1$	0.1	0.02	0.02	0.03	0.02	0.08	0.2	0.08	0.2
$\theta_2$	0.15	0.01	0.02	0.01	0.01	0.09	0.3	0.14	0.2
$\sigma_z^2$	10	9.24	1.17	6.78	0.70	6.16	0.58	5.69	0.49
$\sigma_o^2$	NA	9.27	1.13	6.94	0.71	6.74	0.60	6.29	0.51

**Table 3.6** Summary of posterior parameter inference based on model  $Model_2$ , using different number  $n$  of available observations and informative prior (Table 3.2).

Parameter	True Value	$n = 8$		$n = 16$		$n = 24$		$n = 32$	
		Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
$\beta_1$	0.4	0.54	0.22	0.39	0.11	0.39	0.09	0.42	0.08
$\beta_2$	0.2	0.22	0.06	0.21	0.04	0.2	0.03	0.22	0.02
$\beta_3$	0.01	0.0	0.01	0.01	0.0	0.01	0.0	0.01	0.0
$\theta_1$	0.1	0.06	0.04	0.06	0.02	0.09	0.02	0.08	0.02
$\theta_2$	0.15	0.0	0.02	0.01	0.01	0.11	0.02	0.13	0.02
$\sigma_z^2$	10	2.49	2.57	1.50	1.24	0.84	0.58	1.38	0.85
$\sigma_o^2$	NA	8.17	2.58	8.04	1.44	9.15	0.82	8.62	1.04

**Table 3.7** SD statistics for spatial varying autoregressive parameters  $\rho$ .

SD	$n = 8$	$n = 16$	$n = 24$	$n = 32$
$Model_2$	0.07	0.06	0.02	0.01
$Model_2^*$	0.04	0.09	0.02	0.01
$Model_3$	0.09	-	-	0.008
$Model_3^*$	0.09	-	-	0.008



**Figure 3.3** True and interpolated surfaces of  $\rho$  based on an informative prior and  $n = 32$ . The shading interval is uniform.

The following Tables 3.8, 3.9, 3.10 and 3.11 summarize the results of parameter inference for models  $Model_3$  and  $Model_4$ . For computational reasons we restrict consideration in all analysis to datasets  $n = 8$  and  $n = 32$ . The results suggest similar conclusions about model parameters that have been made for the models  $Model_1$  and  $Model_2$

**Table 3.8** Summary of posterior parameter inference based on model  $Model_3$ , using different numbers  $n$  of available observations and noninformative prior (Table 3.1).

Parameter	True Value	$n = 8$		$n = 32$	
		Mean	s.d.	Mean	s.d.
$\beta_1$	0.4	0.78	0.23	0.48	0.08
$\beta_2$	0.2	0.19	0.07	0.19	0.03
$\beta_3$	0.01	0.01	0.01	0.01	0.0
$\theta_1$	0.1	0.03	0.01	0.12	0.02
$\theta_2$	0.15	0.02	0.01	0.15	0.01
$\phi_r$		(-5.2,-0.42,1.0)	(2.44,0.43,0.23)	(0.07,-0.3,1.0)	(0.88,0.47,0.22)
$\phi_c$		(1.92,-0.51,1)	(1.93,0.5,0.22)	(1.28,-0.27,1.0)	(1.45,0.44,0.24)
$\sigma_z^2$		8.45	1.05	5.33	0.48
$\sigma_o^2$		8.65	1.1	6.76	0.52
$(\sigma_r^2, \sigma_{r0}^2)$		(0.11, 0.07)	(0.16,0.09)	(0.1,0.08)	(0.14, 0.09)
$(\sigma_c^2, \sigma_{c0}^2)$		(0.08, 0.08)	(0.1,0.1)	(0.09,0.08)	(0.12, 0.1)
$\sigma_b^2$		1	1.48	1	0.52

**Table 3.9** Summary of posterior parameter inference based on model  $Model_3$ , using different numbers  $n$  of available observations and informative prior (Table 3.2).

Parameter	True Value	$n = 8$		$n = 32$	
		Mean	s.d.	Mean	s.d.
$\beta_1$	0.4	0.7	0.17	0.48	0.07
$\beta_2$	0.2	0.16	0.06	0.18	0.02
$\beta_3$	0.01	0.0	0.01	0.01	0.0
$\theta_1$	0.1	0.09	0.03	0.11	0.01
$\theta_2$	0.15	0.08	0.04	0.15	0.01
$\phi_r$		(-0.15,-0.26, 1)	(1.21,0.46,0.24)	(0.12,-0.31,1)	(1.04,0.46,0.23)
$\phi_c$		(1.16,-0.35, 1)	(1.38,0.45,0.22)	(1.86,-0.25,1)	(1.85,0.46,0.24)
$\sigma_z^2$		0.83	0.64	0.56	0.29
$\sigma_o^2$		8.12	1.11	9.36	0.56
$(\sigma_r^2, \sigma_{r0}^2)$		0.08, 0.07	(0.11, 0.09)	(0.1,0.08)	(0.17,0.09)
$(\sigma_c^2, \sigma_{c0}^2)$		0.09,0.08	(0.17,0.1)	(0.08,0.08)	(0.09,0.1)
$\sigma_b^2$		1	1.32	1	0.58

**Table 3.10** Summary of posterior parameter inference based on model  $Model_4$ , using different numbers  $n$  of available observations and noninformative prior (Table 3.1).

Parameter	True Value	$n = 8$		$n = 32$	
		Mean	s.d.	Mean	s.d.
$\rho$	0.4	0.44	0.05	0.45	0.02
$\beta_1$	0.4	0.37	0.06	0.4	0.03
$\beta_2$	0.2	0.23	0.03	0.22	0.01
$\beta_3$	0.01	0.01	0.0	0.01	0
$\theta_1$	0.1	0.06	0.02	0.8	0.01
$\theta_2$	0.15	0.12	0.04	0.16	0.01
$\phi_r$		(-0.39,-0.18,1)	(1.2,0.44,0.24)	(-0.3,-0.16,1)	(0.59,0.44,0.23)
$\phi_c$		(-2.5,-0.41,1)	(1.94,0.42,0.24)	(-3.09,-0.61,1)	(3.41,0.43,0.24)
$\sigma_z^2$		8.42	1.07	4.80	0.4
$\sigma_o^2$		9.45	1.26	6.83	0.52

$(\sigma_r^2, \sigma_{r0}^2)$	(0.09,0.08)	(0.12, 0.1)	(0.1,0.08)	(0.15, 0.1)
$(\sigma_c^2, \sigma_{c0}^2)$	(0.1,0.08)	(0.19,0.1)	(0.09,0.08)	(0.08,0.11)
$\sigma_b^2$	2	2.19	1	1.21

**Table 3.11** Summary of posterior parameter inference based on model  $Model_4$ , using different numbers  $n$  of available observations and informative prior (Table 3.2).

Parameter	True Value	$n = 8$		$n = 32$	
		Mean	s.d.	Mean	s.d.
$\rho$	0.4	0.43	0.04	0.42	0.02
$\beta_1$	0.4	0.37	0.04	0.4	0.02
$\beta_2$	0.2	0.22	0.02	0.21	0.01
$\beta_3$	0.01	0.01	0	0.01	0
$\theta_1$	0.1	0.07	0.02	0.08	0.01
$\theta_2$	0.15	0.14	0.03	0.16	0.01
$\phi_r$		(0.01,-0.23,1)	(1.08,0.44,0.24)	(-0.67, -0.11, 1)	(0.71,0.45,0.23)
$\phi_c$		(-1.84,-0.4,1)	(2.02,0.42,0.23)	(-2.13,-0.6,0.9)	(3.86,0.46,0.24)
$\sigma_z^2$	10	0.81	0.62	0.57	0.27
$\sigma_o^2$		9.46	1.26	9.11	0.57
$(\sigma_r^2, \sigma_{r0}^2)$		(0.09, 0.08)	(0.11,0.09)	(0.11,0.08)	(0.09,0.08)
$(\sigma_c^2, \sigma_{c0}^2)$		(0.1, 0.08)	(0.2,0.1)	(0.09,0.08)	(0.1,0.15)
$\sigma_b^2$		2	2.04	1	1.3



### 3.5.2 Model performance

We now investigate the spatial-temporal performance of the proposed models  $Model_1$ ,  $Model_2$ ,  $Model_3$  and  $Model_4$  and how this depends on different datasets and priors. Additionally, the proposed models also allow us to investigate the local transport effect on concentration levels of the true process. To quantify the agreement between observed and estimated concentration levels we calculate the root mean squared error (RMSE) and to assess the relative importance of the transport effect we use the ratio of the variances of the transport term to the total variance.

Table 3.12 reports the model performance and relative importance of the transport effect under different datasets and priors. We note that the average RMSE decreases with an increase in the available observations and, by contrast, the average transport effect increases with an increase in the available observations. Using informative prior specification improves model performance in all cases. This result could be connected to the observed improvement in the parameter inferences when we use informative prior specification and in particular the variance parameters, which are the most sensitive parameters to the prior specification. Another important point is that the spatially varying autoregressive models  $Model_2$  and  $Model_3$  do not perform better than models  $Model_1$  and  $Model_4$  with smaller datasets. The reason for this could be the poor modeling assumption; the way we generate our datasets and/or the natural requirement that more parameters require larger dataset to be estimated. To clarify the latter we performed an additional run of the model  $Model_2$  on the larger dataset ( $n = 8$  and  $T = 100$ ). As a result, we observe improvement of 14% and better in model performance. This shows that these models are not suitable for small datasets.

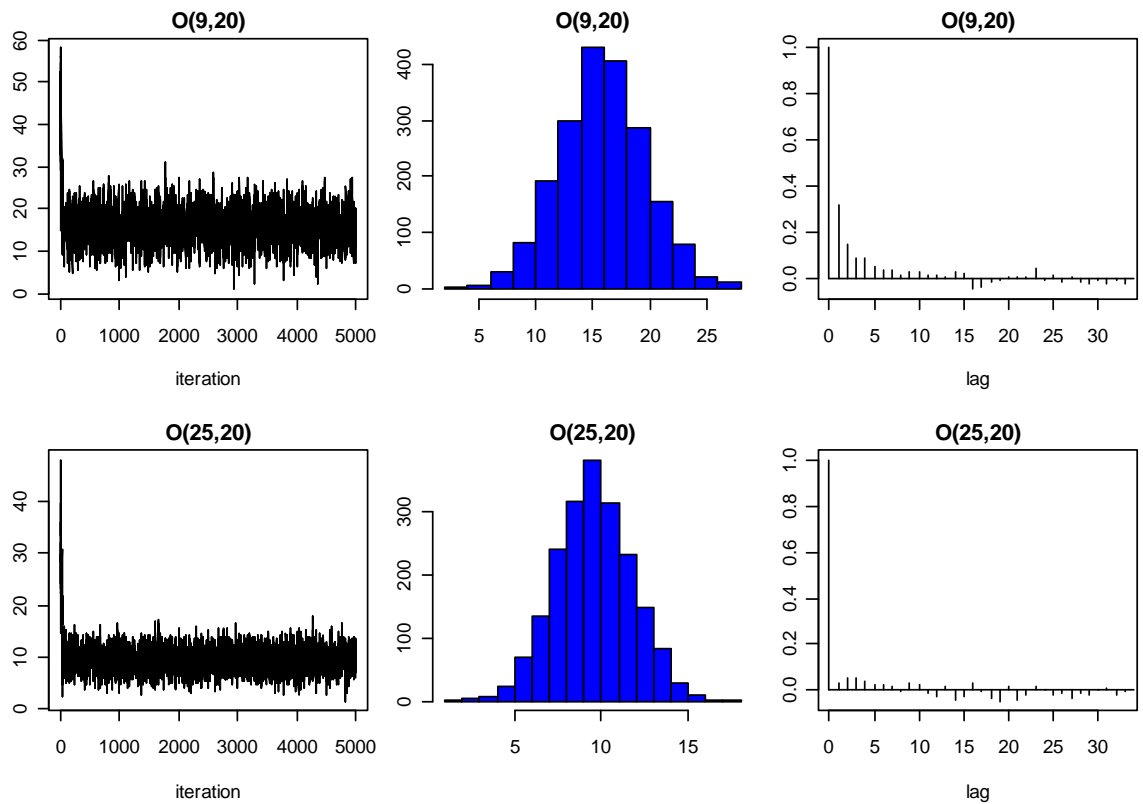
The model performance might eventually improve if we consider some form of spatial interaction among the time autoregressive parameters such as MRF. This option was not investigated here.

**Table 3.12** Summary of the model’s performance and relative importance of the transport effect. \* denotes that informative prior was used for calculations.

Model	$n = 8$		$n = 16$		$n = 24$		$n = 32$	
	RMSE	Tr.	RMSE	Tr.	RMSE	Tr.	RMSE	Tr.
$M_1$	3.50	3.7%	3.19	5.4%	2.79	7.5%	2.34	9%
$M_1^*$	3.35	10.3%	2.94	11.7%	2.36	16%	1.7	17.6%
$M_2$	14.65	2.4%	13.95	3.4%	3.64	9.6%	2.89	10.9%
$M_2^*$	9.43	7.2%	5.67	16.2%	3.34	22%	2.57	16.3%
$M_3$	19.99	3.4%	-	-	-	-	2.55	11.2%
$M_3^*$	17.27	16.5%	-	-	-	-	1.96	22%
$M_4$	3.34	8%	-	-	-	-	2.23	9%
$M_4^*$	3.2	12.9%	-	-	-	-	1.61	16.9

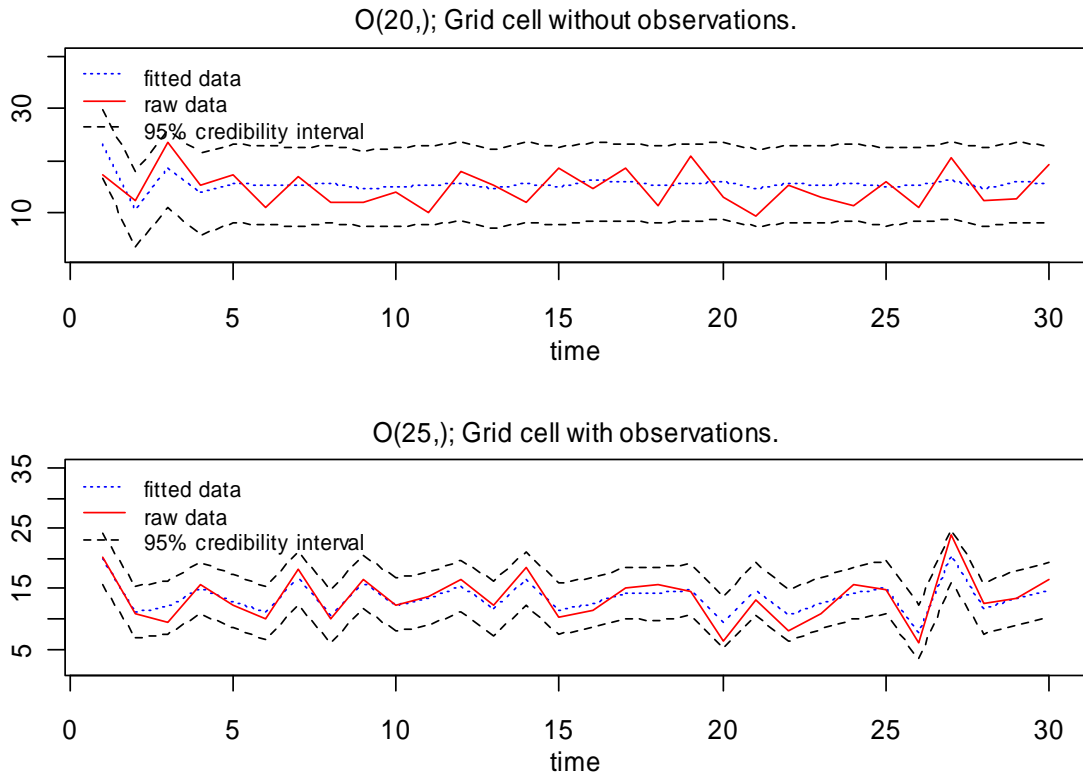
Figure 3.4 (center) presents MCMC histograms of the response process from randomly selected grid cells with observations (grid cell number 25) and without observations (grid cell number 9), based on model  $Model_1$  ( $n = 8$ , noninformative prior) at time point  $t = 20$ . The posterior distribution of the grid cells that do not have observations is wider indicating that much more uncertainty is associated with response estimates where no data are presented. Furthermore, in this case the autocorrelation (right panel) is higher which implies slow convergence of the Gibbs Sampler.

The local impact of data availability in time is further investigated in Figure 3.5 where the posterior time series of the true processes are presented for the same locations. The 95% credibility lines represent the evolution of the uncertainty in time. It is clear that the credibility interval is wider at each time point for the grid cell that has no observations within the grid cell. Another point is that the 95% credibility interval for the grid cell 9 is relatively constant in time and does not follow closely the time variations of the true process (the solid line). The posterior time series for grid cell 9 also fails to



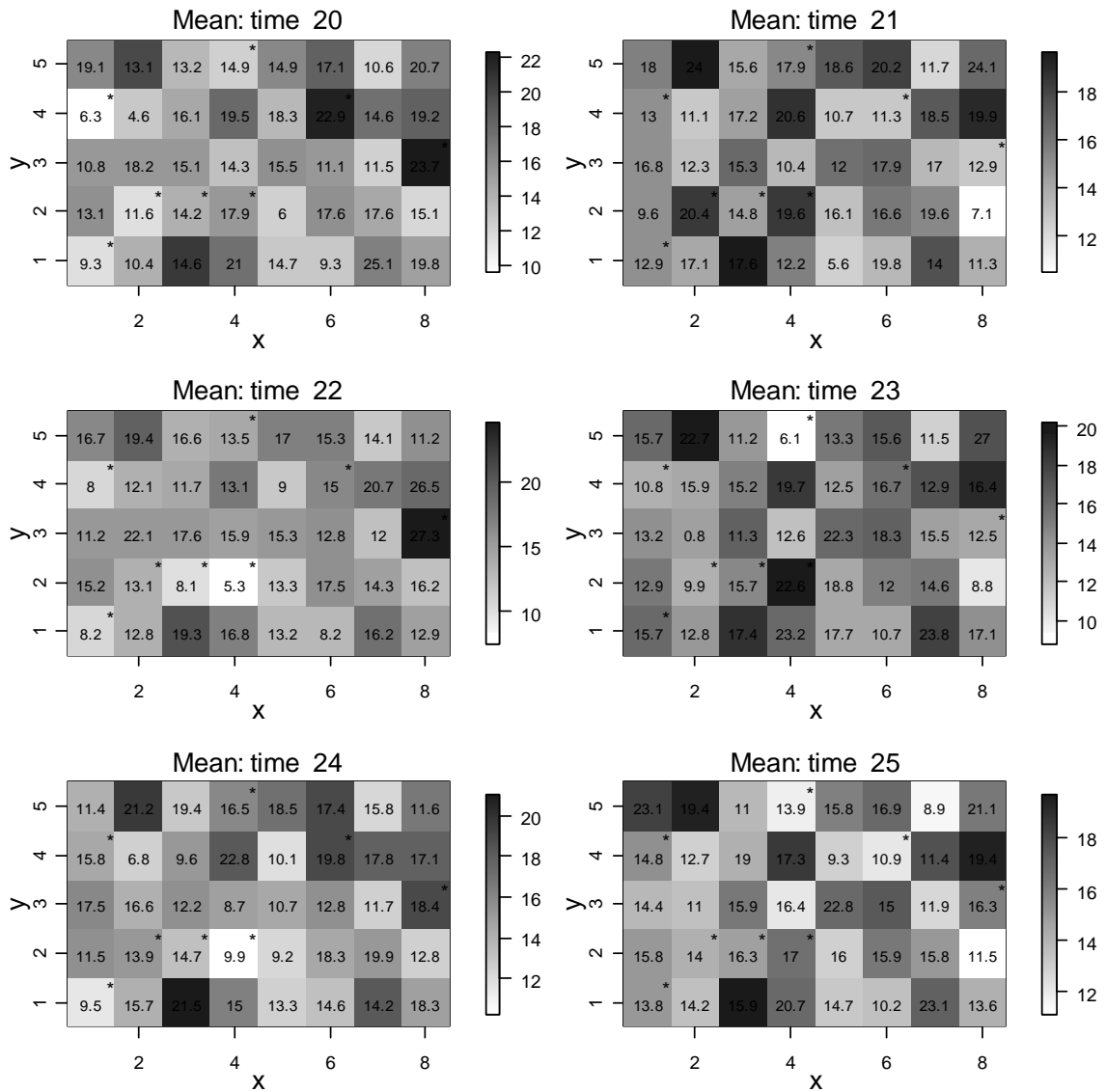
**Figure 3.4** Left: sequences of 5000 iterations from the posterior distributions of the randomly selected grid cells without observations  $O(9,20)$  and with observations  $O(25,20)$  ; Center: histogram of the last 2000 iterations; Right: autocorrelation function.

capture most part of the variability of the true process even though on average, these two time series are very close. By contrast, the posterior time series for grid cell 25 that has observations within the grid cell follows very closely the observed time series as well as the 95% credibility interval. We also should note that the credibility interval is narrow in this case and in both cases the observed time series is within the 95% credibility interval.



**Figure 3.5** Posterior mean time series for grid cells without observations  $O(9,\cdot)$  and for grid cells with observation  $O(25,\cdot)$ .

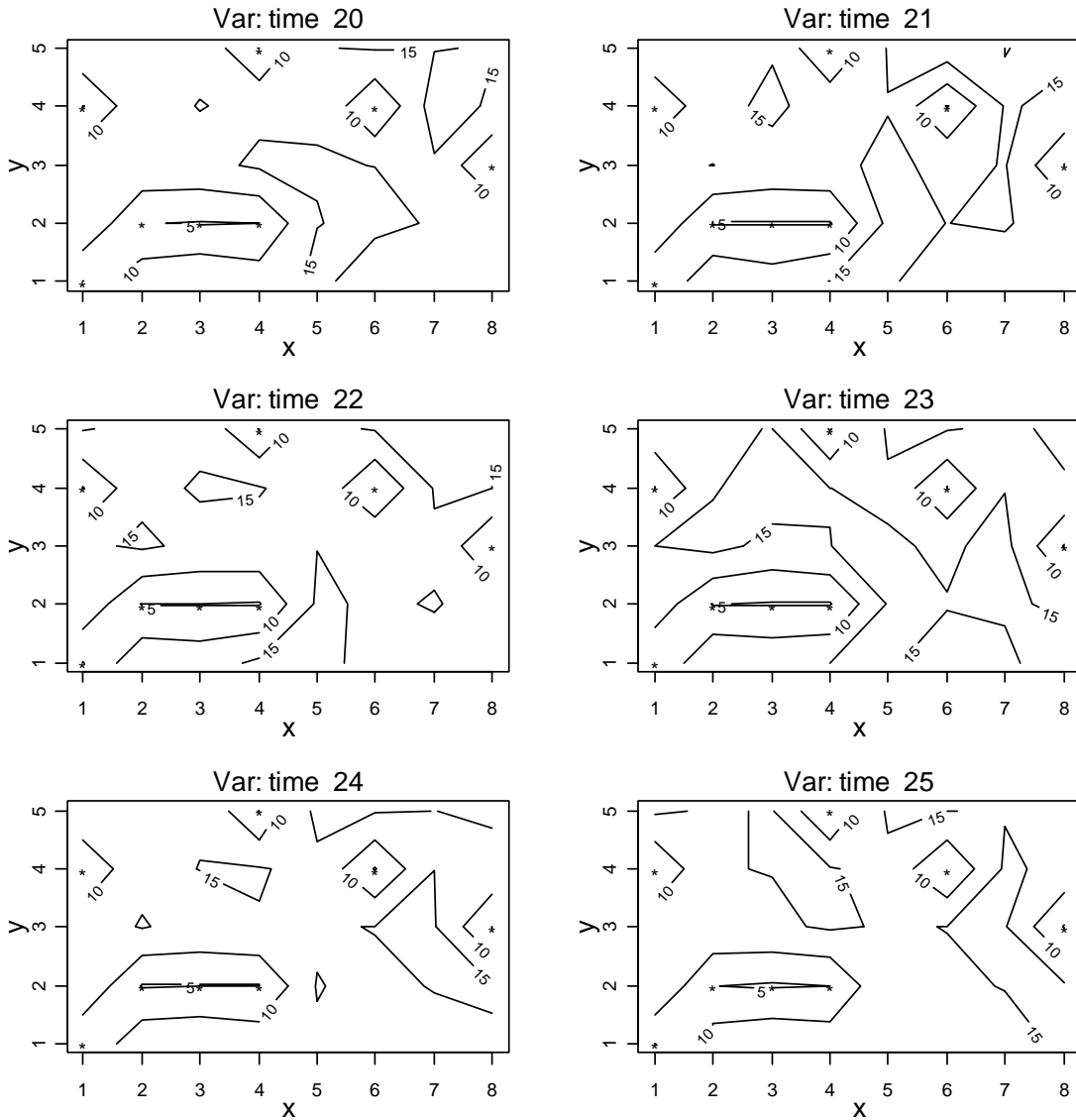
Model based interpolated maps of the true process (Figure 3.6) are given for the randomly selected time period (time interval 20 to 25). We found good agreement with the observations superimposed on the plots and that the model was able to capture correctly the space-time dynamic evolution. The RMSEs for the period are 3.55, 3.04, 2.70, 3.80, 3.05, and 2.78 respectively. Similar results were obtained for the other datasets  $\{16, 24, 32\}$  and models  $\{Model_2, Model_3, Model_4\}$  (not shown). The average RMSE for the whole period of  $T = 30$  equally spaced time points are reported in Table 3.12.



**Figure 3.6** Model-estimated mean field of the true pollutant concentration (grey scale), and observed values (numbers). \* - denotes the grid cells used for fitting the model.

Figure 3.7 shows the corresponding standard deviation contour plots. The plots illustrate, as expected, that the uncertainty is lowest near the grid cells that have observations. Since the standard deviation is very sensitive to the data availability these plots are

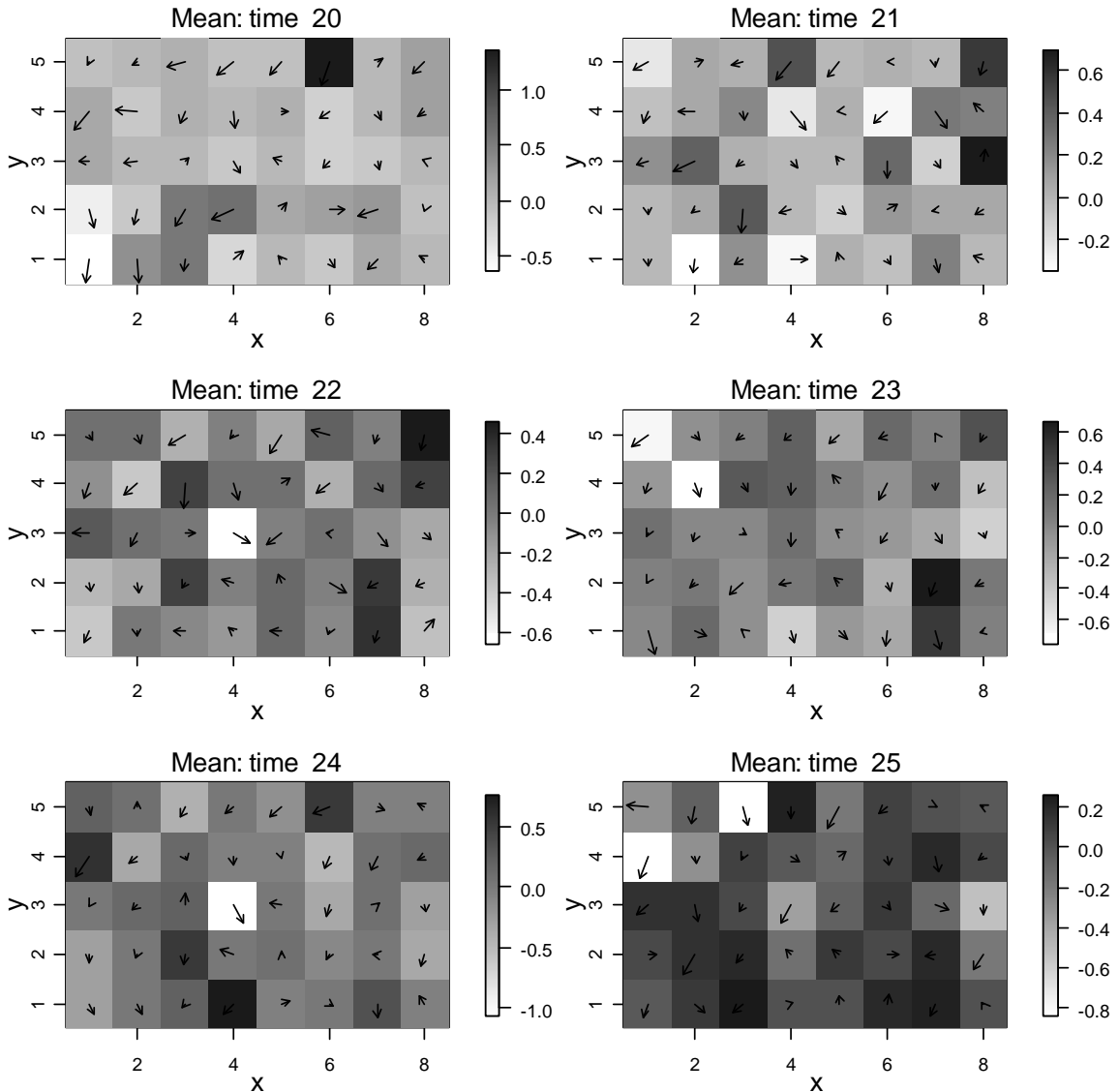
not very suitable for characterizing the domain variability. Large-scale variability patterns might be very different from those presented in Figure 3.7.



**Figure 3.7** Contour plots of pollutant variance field. \* - denotes the grid cells used for fitting the model.

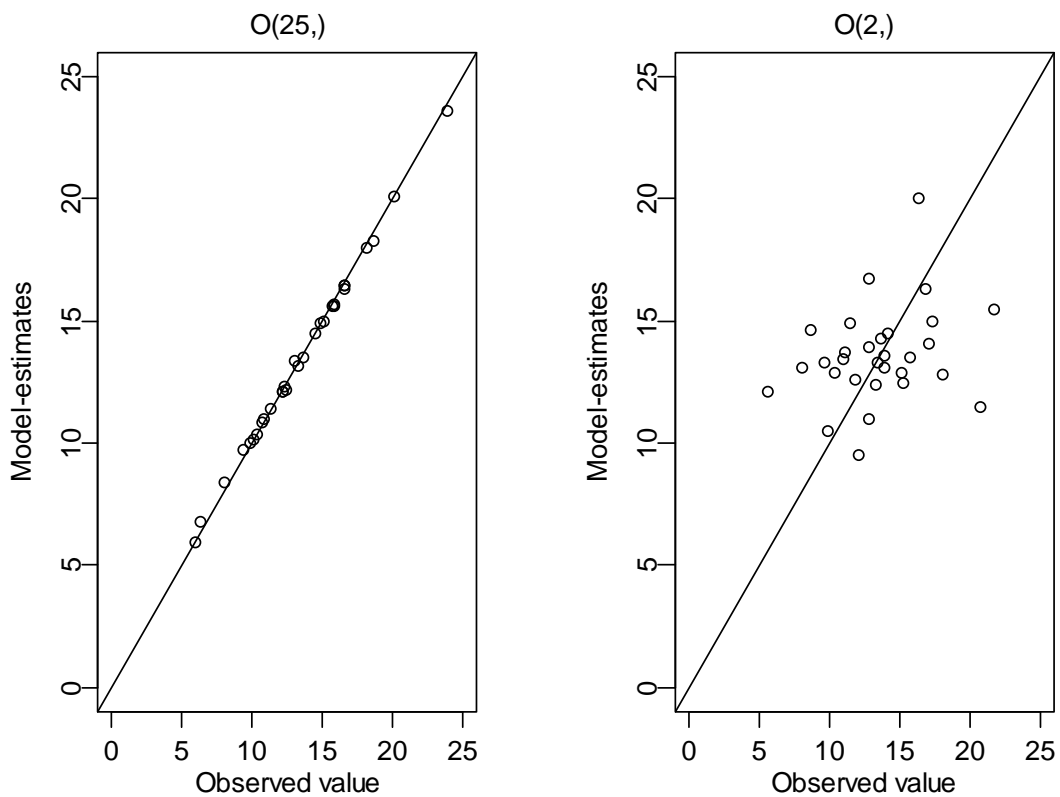
The model's ability to estimate the transport effect is demonstrated in Figure 3.8. In general the effect is small and the wind can be associated with positive, negative and

no local transport contributions to the pollutant concentration levels. The sign of the transport effect is determined by the pollutant concentration gradients and the magnitude of the effect depends on wind speed and wind direction as well magnitude of gradients. Again, we do not try to find large-scale transport patterns since the data are randomly generated. The average overall transport effect is reported in Table 3.12.



**Figure 3.8** Model estimated mean transport contributions map. . The observed wind field is superimposed on the plots.

We have also performed the usual diagnostic plots for checking model adequacy. Figure 3.9 shows the scatter plot of the estimated values against observed values for the two different types of grid cells. These results agree with our previous results and conclusions reported in Table 3.12 and Figure 3.5. We have observed similar behavior for all datasets and models used in this simulation study.



**Figure 3.9** Scatter plot for grid cell with observations  $O(25, \cdot)$  and for grid cell without observation  $O(2, \cdot)$ .



### 3.6 Discussion

In this simulation study we investigated different aspects of the general model proposed in Chapter II including: how we treat the boundary sites; the way we incorporate uncertainty by allowing spatial varying autoregressive parameters or not; the data availability; and prior specification. We found that:

- In general the model parameters can be divided into three groups independent of differences in model specification. In the first group we have well estimated parameters independent of data availability and prior specification, such as  $\beta$ . In the second group we have parameters whose estimates can be substantially improved depending on data availability and prior specification, such as  $\theta$ , and  $\sigma^2$ . The third group includes parameters that are not very sensitive to prior specification or data availability, such as  $\rho$ .
- Model performance and the transport effect depend significantly on spatial data availability (the ratio of the number of the cells without observations to the number of the cells with observations, i.e.  $n/N$ ) and prior specification especially for the models that allow spatial varying autoregressive parameters, *Model*<sub>2</sub>, and *Model*<sub>3</sub>. The reason for the relatively poor performance of some of the models could be the length of the data records; poor model assumptions; or the way we generate the datasets. Therefore, it would be worthwhile to investigate further simulation studies with larger datasets and other ways for incorporating uncertainty.
- There is a considerable difference between grid cells that have or do not have observations within the cells. The grid cells that do not have observations are associated with much more uncertainty. This would explain the strong dependence of model performance on spatial data availability. Increasing the number

of the cells with observations will decrease the overall uncertainty and vice versa, and ultimately model performance.

- The Gibbs Sampler slows significantly as the complexity of the model increases.

Additional runs of the models have been carried out to investigate the sensitivity of the results on deliberately misleading prior specifications. That is, we incorrectly specified the prior information for different groups of model parameters  $(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\sigma})$  in factor of 3-4. In all cases the result was very similar to the one obtained using correctly specified priors.

## **CHAPTER 4**

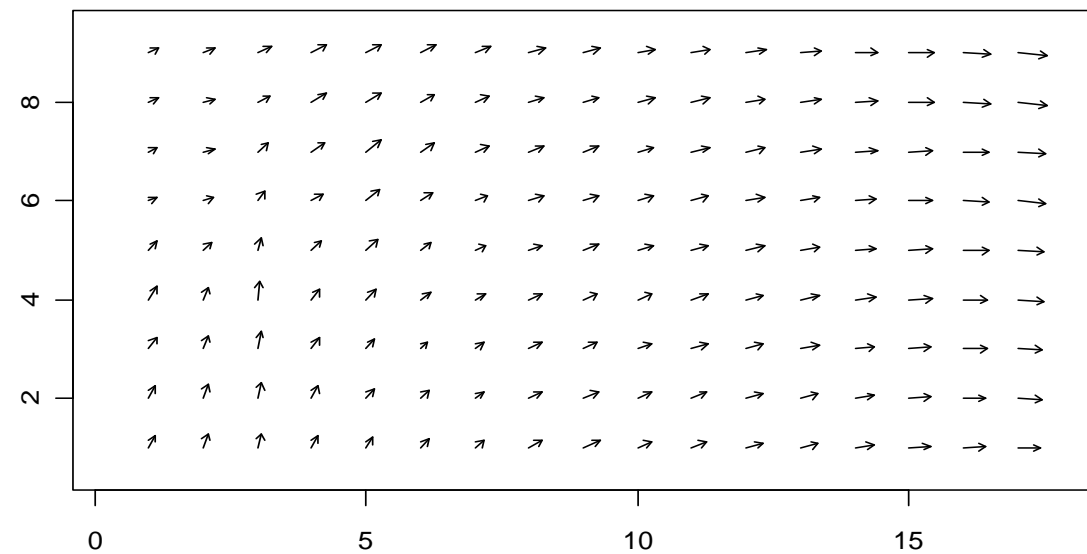
### **APPLICATION TO GROUND LEVEL OZONE CONCENTRATION DATA**

#### **4.1 The Metro Vancouver data**

The dataset considered here consists of one-hour-average ozone measurements provided the Lower Fraser Valley (LFV) Air Quality Monitoring Network, covering the time period of 24 years from 1982 to 2006 at 38 stations (Lower Fraser Valley Ambient Air Quality Report, 2007). However, we model daily maximum 8-hour average ozone concentration data obtained from hourly ozone data by calculating the maximum of rolling 8-hour averages for each day. The 8-hour average values are of particular interest because that is the averaging basis for the Canadian-Wide Standards for ozone, based on the adverse effect of ozone and other pollutants on human health and the environment (EPA AQCD, 2006). The spatial domain of study is restricted to a rectangular region with size of about 146x61 km<sup>2</sup> covering the Lower Fraser Valley. We also restrict the time period of analysis the most recent 3 years available (2004, 2005, and 2006). That period includes one year of high concentrations (2006), one of low (2005) and one of intermediate (2004), though the analysis of data from other years is not excluded. The ozone season considered here starts from 15 April through 30 September, 169 days in all (McMillan et al., 2004). All hourly measurements have a quality control flag and in this analysis we use only high quality data from stations with at least 90% data availability. These criteria should optimize model spatial interpolation performance and partially eliminate any uncertainty in our results that might be attributed to measurement error or more generally inadequate data quality.

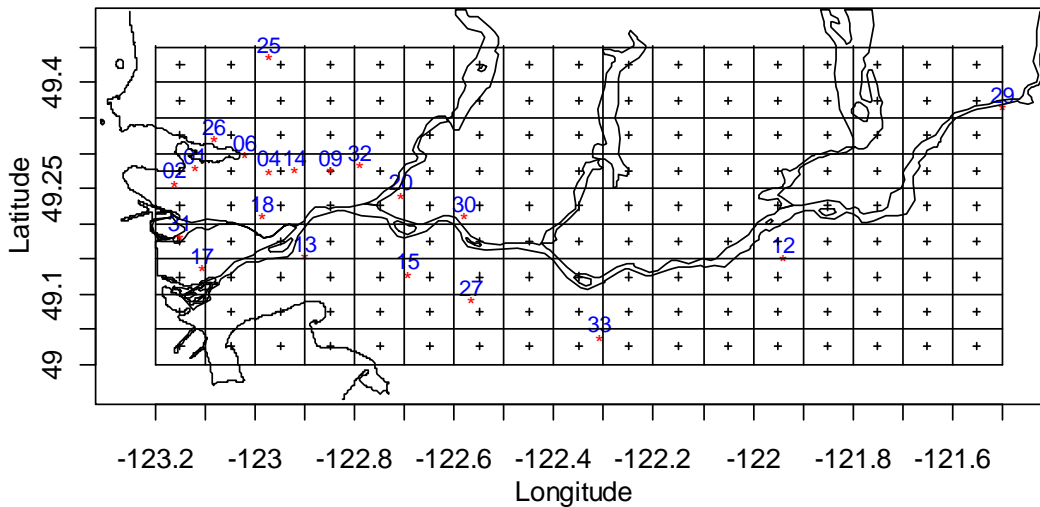
In addition, we have meteorological data from 13 weather stations located at some ozone monitoring sites, since our modeling approach requires meteorological fields as covariates. We interpolate the covariates fields to a regular grid specified over the entire domain of study for each time point using ordinary kriging. Once we obtain the covariate fields we calculate the daily rolling 8-hour averages of meteorological covariates at each grid cell, the daily 8-hour average values that are used as input in analysis. For grid cells that have no ozone observations within we select values that correspond to the domain daily average maximum 8-hour ozone concentration. Figure 4.1 displays the resulting interpolated 8-hour average wind field for 12 May, 2004. Alternatively, we can use the output of a Numerical Weather Prediction (NWP) system to obtain the meteorological fields.

We should also note that, first: the wind field is a vector variable described by wind speed and direction that was decomposed into two orthogonal components and then interpolated over the grid (Stull, 1988); second: not all stations were operational at the same time; third: all of the data points have a quality control flag that indicates possible problems that we do not believe affect the conclusions drawn from our analysis.



**Figure 4.1** Interpolated 8-hour average wind field for 12 May 2004

Figure 4.2 provides a geographical summary of the spatial domain and the location of the 20 air quality sites that had available data for the time period of study. The figure also shows the region over which we define the interpolation grid for applying the spatiotemporal model and that the spatial data coverage is approximately 11%.



**Figure 4.2** Locations of ozone observation sites “\*” and prediction grid “+”. Numbers are station identifiers.

Different aspects of ozone behavior in the LFV have been studied by many authors including: McKendry, 2002; Le and Zidek, 2006; Robeson and Steyn, 1989, 1990; Pryor et. al, 1994; Pryor and Steyn, 1995; Joe et. al, 1996; Vingarzan and Taylor, 2003; Burrows et. al, 1995; Ainslie and Steyn, 2007. The range of methods used in these studies vary from simple time series and regression models to complex hierarchical Bayesian models and Artificial Neural Networks.

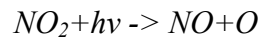
Restricting the time domain to a particular season is appropriate for the time scale of an operational model and was done in several previous studies: McMillan et al., 2005;

Huang and Hsu, 2004; Huerta et al., 2004; Wikle and Cressie, 1999. In addition, it keeps the problem computationally tractable and facilitates the modeling approach since the seasonal and longer scale cycles appear as a linear trend.

## 4.2 Background chemistry of ozone

Different aspect of ozone as an air pollutant have been studied: hourly surface level concentrations (e.g. Gardner and Dorling, 2000), daily maximum concentration (e.g. Burrows et. al, 1995), annual cycle (e.g. Vingarzan and Taylor, 2003), threshold exceedance (e.g. Smith and Huang, 1993), correlated variables (e.g. McKendry, 2002), evaluation of an air quality model (e.g. Riccio et. al, 2005), systematically missing data (e.g. Le and Zidek, 2006), and trends assessment (e.g. Rao and Zurbenco, 1994, Sahu et al., 2007).

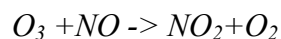
Ozone is secondary pollutant which forms by the photochemical destruction of nitrogen dioxide ( $\text{NO}_2$ ) in the presence of sunlight. As a result of the reaction, nitric oxide (NO) and an atom of oxygen (O) are created.



Here the energy of the solar radiation is presented by the second term;  $h\nu$ , which is the product of Planck's constant,  $h$ , and the frequency,  $\nu$ , of solar electromagnetic radiation. Next, the oxygen atom combines with molecular oxygen ( $\text{O}_2$ ) in the presence of a third molecule that absorbs the excess energy to produce ozone ( $\text{O}_3$ )



In the absence of volatile organic compounds (VOCs) ozone reacts with nitric oxide to reform nitrogen dioxide



The presence of VOCs and CO amplifies ozone production by removing NO and creating additional NO<sub>2</sub>. When sufficient amounts of NO<sub>2</sub> and VOCs are present in appropriate weather conditions, ozone formation exceeds dissociation, causing it to build up in the atmosphere. Ozone is also transported to the troposphere from the stratosphere.

The sources of tropospheric zone precursors, NO<sub>x</sub> and VOCs, in general can be grouped in two major categories namely: biogenic (natural) and anthropogenic. The natural NO<sub>x</sub> sources include stratospheric intrusions, lightning, soils, and wildfires and the two largest anthropogenic NO<sub>x</sub> sources are electric power generation plants and motor vehicles. Uncertainties in natural NO<sub>x</sub> emissions are much larger than for anthropogenic NO<sub>x</sub> emissions. Naturally emitted VOCs are produced by vegetations and the largest anthropogenic sources are industrial processes and transportation. On the regional and global scales, emissions of VOCs from vegetation are much larger than those from anthropogenic sources. On an urban scale, uncertainties in both biogenic and anthropogenic VOCs emission inventories prevent determination of the relative contributions of these two categories with much precision (EPA AQCD, 2006).

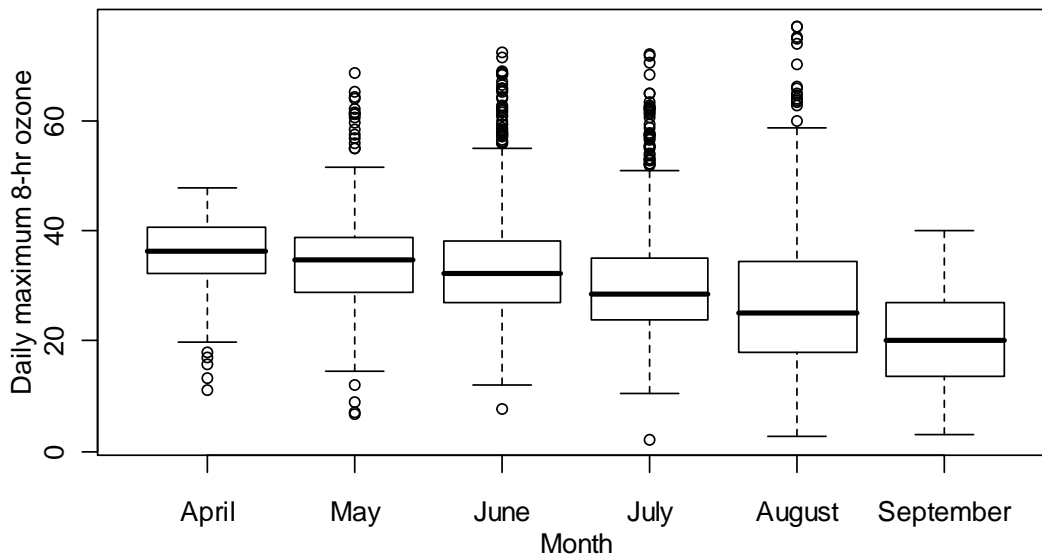
Ozone may exhibit significant temporal variations on different scales, for example, seasonal cycles caused by increases in biogenic emissions during the summer when the temperature and incident sunlight are highest. Temperature and sunlight are also responsible for an ozone diurnal cycle, while climate changes are responsible for long term trends (EPA AQCD, 2006).

Additionally, ozone may exhibit significant spatial variation as well. For example, in urban areas, it is well known that ozone concentration is lowest in the center and increases in the downwind direction. The reason is that efficiency of ozone production per NO<sub>x</sub> oxidized is highest in areas where NO<sub>x</sub> concentrations are lowest (suburban/rural areas) and decreases with increasing NO<sub>x</sub> concentration (urban centers). The atmospheric chemistry of tropospheric ozone is a complex, non-linear function of many factors, which

goes through several stages. A comprehensive overview of the problem is given in Seinfeld and Pandis, 1998.

### 4.3 Exploratory Data Analysis (EDA)

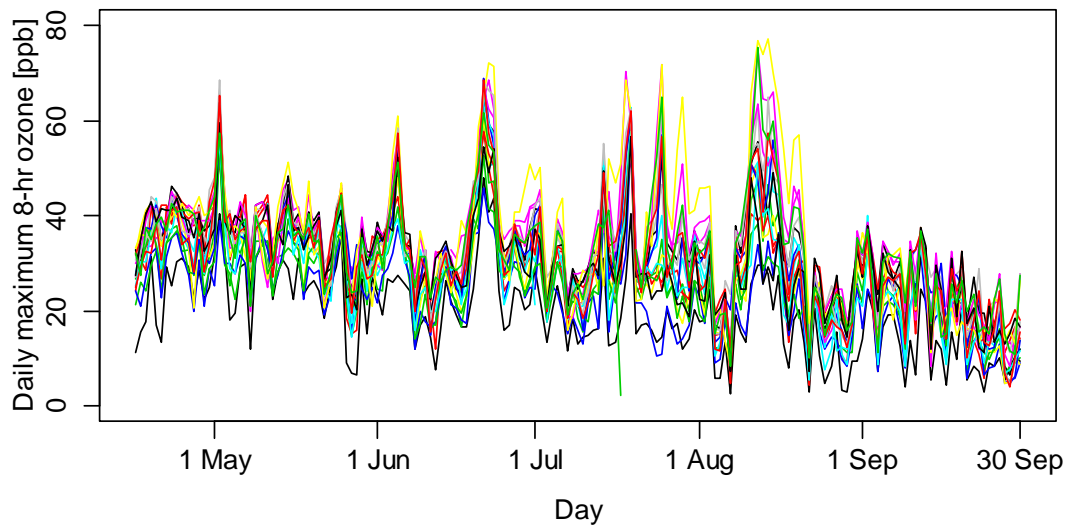
In order to gain some understanding of the spatial and temporal patterns presented in the data, we have conducted a preliminary EDA. Figure 4.3 presents boxplots of the domain daily maximum 8-hour ozone values by month. It is clear that there is a temporal trend over the period of study. The domain-average levels steadily decrease from the beginning of the period (April 15) to the end of the period (September 30).



**Figure 4.3** Boxplot of the domain daily maximum 8-hour ozone levels by month.



This decreasing trend is also seen in the daily maximum 8-hour ozone time series (Figure 4.4), suggesting that a simple linear trend would be a reasonable large-scale temporal model,  $\alpha(t)$ , for the data in this case.

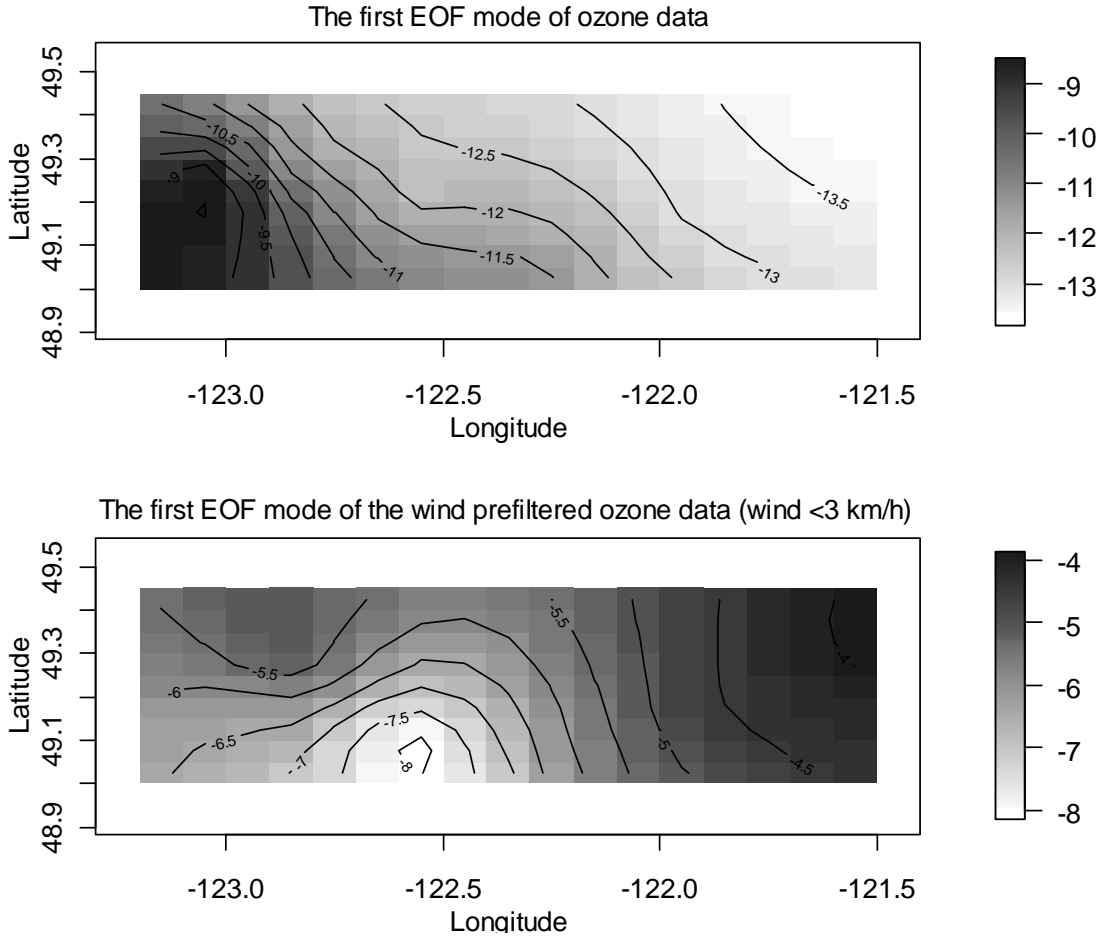


**Figure 4.4** Time series of observed daily maximum 8-hour ozone concentrations at 20 monitoring sites from April 15th to September 30th.

A Principal Component Analysis (PCA) was employed to investigate the large-scale spatial features that are mostly dominated by the spatial variability of emissions rates of ozone precursors throughout the domain of interest. Since EOFs are defined to be orthogonal and the PCs are defined to be uncorrelated they are an efficient means of isolating coherent modes of variability in large data sets (McPhaden and Hayes, 1991). To perform PCA we use an iterative Bayesian model that is able to handle the missing values problem. The model is implemented in an R package namely: `pcaMethods`, (Shigeyuki et al., 2003).

The top panel of Figure 4.5 shows the first empirical orthogonal function (EOF) of ozone data. This spatial pattern explains 85.7% of the total variance and is similar in structure to the mean ozone field (not shown). The bottom panel of Figure 4.5 plots the first EOF mode of wind prefiltered ozone data. By this we mean we used only ozone values that correspond to very light winds less than or equal to 3 km/h; similar results were obtained for 2 km/h. In addition, this eigenfunction describes 68.37% of the total variance and clearly indicates the considerable impact caused by the transport effect on the mean field.

Both panels in Figure 4.5 present a complex large-scale spatial pattern which is very difficult to reproduce by a continuous function of some covariates, e.g. longitude and latitude (Wikle 1998). That is why we choose to use the first EOF mode of wind prefiltered ozone data to model the large-scale spatial trend,  $\beta(s)$ , as was discussed in Chapter 2.



**Figure 4.5** The first EOF modes of ozone data.

Now, we have sufficient information to specify the larger-scale spatiotemporal model (2.4), which in our case takes the form

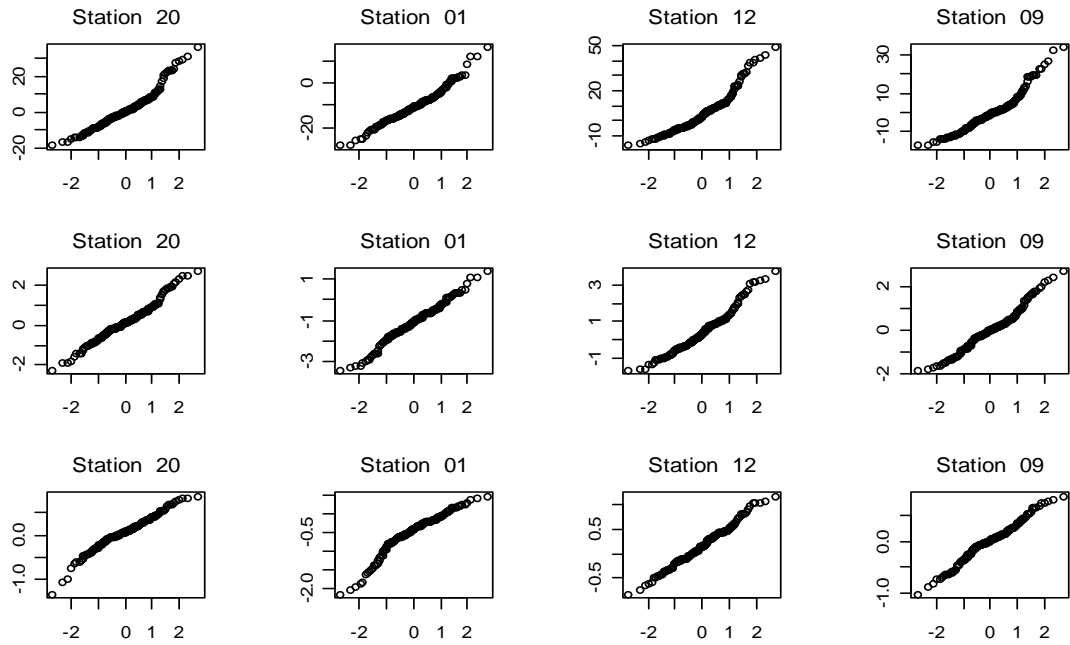
$$\boldsymbol{\mu}_t = (\alpha_1 + \alpha_2 t) \otimes \mathbf{1}_{N \times 1} + \gamma \mathbf{EOF}_1; \quad t = 1, \dots, T. \quad (4.1)$$

Least-square (LS) methods were used for inferring the parameters  $\{\alpha_1, \alpha_2, \gamma\}$ , with estimated values  $\{5.50, -0.01, 0.28\}$ . After removing the large-scale signals from the observations the residuals  $Z(s_i, t) - \hat{\mu}(s_i, t)$ ,  $i = 1, \dots, n$ ;  $t = 1 \dots T$ , where  $n$  is the number

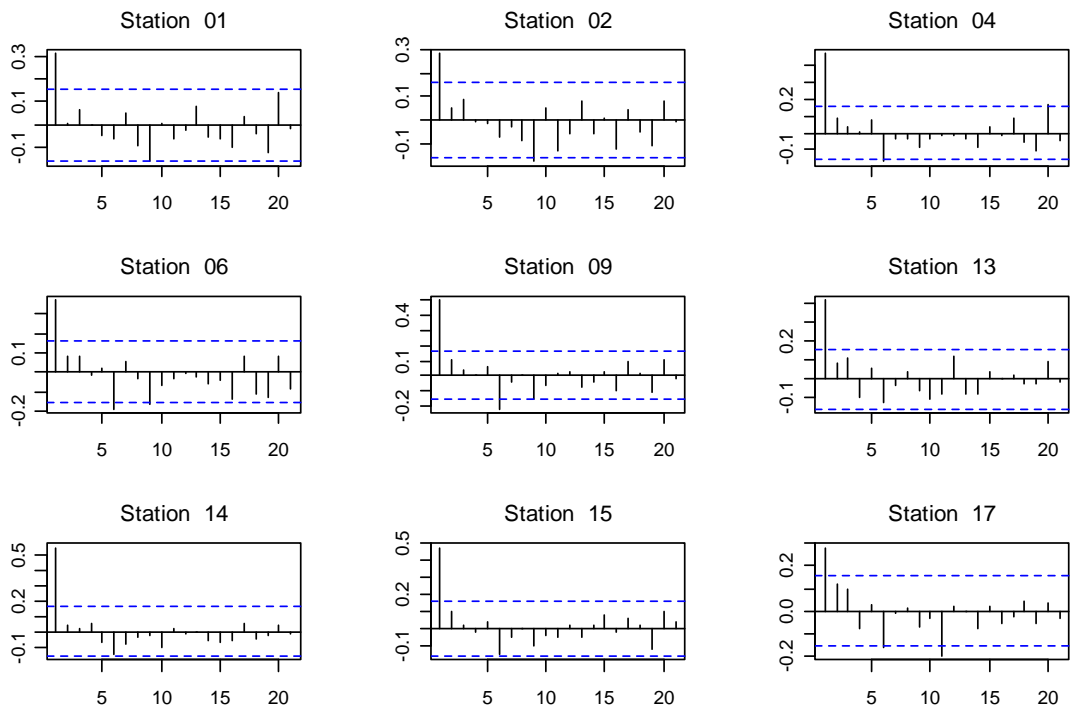
of ozone monitoring sites, were used for fitting the Bayesian hierarchical models  $\{Model_1, Model_2, Model_3, \text{ and } Model_4\}$  specified in Chapter 3.

We investigate the normality of the residuals using histograms and normal quantile-quantile (Q-Q) plots on three measurement scales: original, square root, and logarithmic. Figure 4.6 displays the normal Q-Q plots of four randomly selected ozone monitoring sites on the three scales. It is easy to see that in all cases the normality assumption is reasonable and there is a slight improvement by using variance stabilizing scales (square-root and logarithmic), seen in the middle and bottom panels. However, in some cases, e.g. station 1, the logarithmic transformation appears to have over compensated for the skewness. The histogram plots (not shown here) confirm this observation. This makes us believe that the square-root scale would be the most appropriate scale to work with on in our analysis in prospect of stabilizing the variance and symmetry. This result is in agreement with other works in modeling ozone concentration levels; see for example, Sahu et al., 2007; Carroll et al., 1997; Huerta et al., 2004; Huang and Hsu, 2004.

The partial autocorrelation function for the residuals of the first nine (ordered on station identifiers) time series are plotted on Figure 4.7. The result indicates clearly the cut off at lag one of the partial autocorrelation function that suggest the first order AR process would be appropriate for each station. This result is in accord with the statistical model suggested in Chapter 2, and built on physical arguments. In the suggested model, the time autoregressive term appears naturally from the advection equation.



**Figure 4.6** Normal Q-Q plot of residuals on original scale (top panels), square root scale (middle panels), and logarithmic scale (bottom panels).



**Figure 4.7** Partial autocorrelation functions of the first nine time series.

## 4.4 Results

### 4.4.1 Parameter inference

A Gibbs sampler was applied to estimate the marginal posterior distributions for model parameters. A summary of the results from the Gibbs sampler are presented in Table 4.1. The results suggest the following conclusions:

- The estimates of the regression parameters  $(\beta_1, \beta_2, \beta_3)$  are similar for all models. The temperature ( $\beta_1$ ) is the meteorological parameter that has the most significant influence on average grid cell ozone concentration.
- The autoregressive process for models *Model*<sub>1</sub>, and *Model*<sub>4</sub>,  $\rho$ , has the second most significant influence on ozone levels. Additionally, the spatially varying autoregressive processes for models *Model*<sub>3</sub> and *Model*<sub>4</sub> are shown in Figure 4.8. A visual inspection of these two surfaces shows that they have substantial spatial variation, which supports the use of a model with spatially varying autoregressive coefficients.
- The spatial autoregressive parameters  $(\theta_1, \theta_2)$  have only a minor effect on ozone concentration in all cases. However, the zonal (east-west) component  $\theta_1$  is much bigger than the meridional (north-south) component  $\theta_2$  for models *Model*<sub>2</sub> and *Model*<sub>3</sub> which is in agreement with the prevailing westerly summertime winds in LFV (Ainslie and Steyn, 2007)
- The estimates of the variance components  $\sigma_z^2$  and  $\sigma_o^2$  show that more variance is explained by the second stage error process ( $\sigma_o^2$ ) than by the data error process ( $\sigma_z^2$ ), except *Model*<sub>1</sub>.

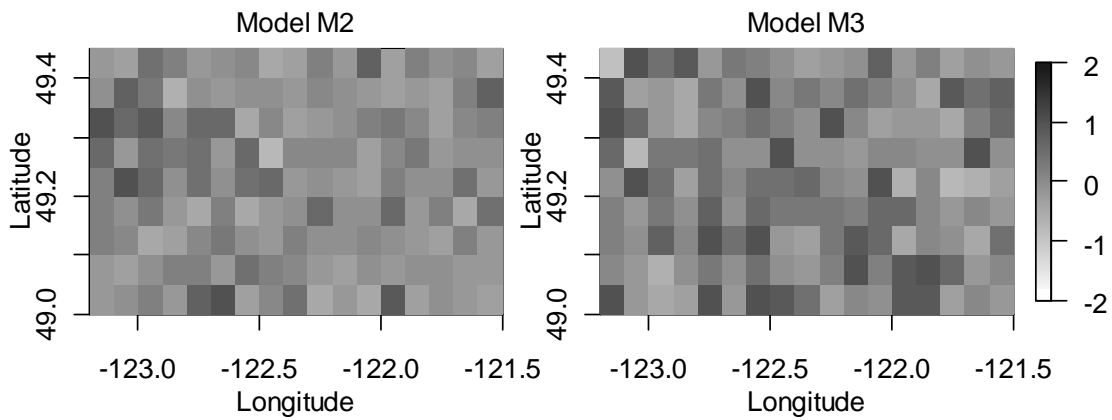
Parameter inferences related to the boundary process that is included in models *Model*<sub>3</sub>, and *Model*<sub>4</sub> are reported in Table 4.2.

**Table 4.1** Summary of posterior parameter inference.

Parameter		<i>Model</i> <sub>1</sub>		<i>Model</i> <sub>2</sub>		<i>Model</i> <sub>3</sub>		<i>Model</i> <sub>4</sub>	
		Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
$\beta_1$	Temperature	0.429	0.031	0.604	0.03	0.586	0.032	0.585	0.034
$\beta_2$	Rel hum	-0.135	0.005	-0.168	0.004	-0.165	0.005	-0.166	0.006
$\beta_3$	Solar rad	0.001	0.001	-0.002	0.001	-0.003	0.001	-0.003	0.001
$\theta_1$	E-W space	-0.006	0.005	0.013	0.001	0.01	0.002	-0.004	0.003
$\theta_2$	N-S space	-0.001	0.006	-0.016	0.003	-0.013	0.002	0.001	0.005
$\rho$	Auto reg.	0.48	0.02					0.479	0.016
$\sigma_z^2$	Data var.	33.33	1.133	20.26	3.137	25.51	2.68	24.86	6.116
$\sigma_o^2$	Process var.	24.1	1.51	25.63	1.53	26.73	1.46	30.61	1.503

**Table 4.2.** Estimation of boundary parameter.

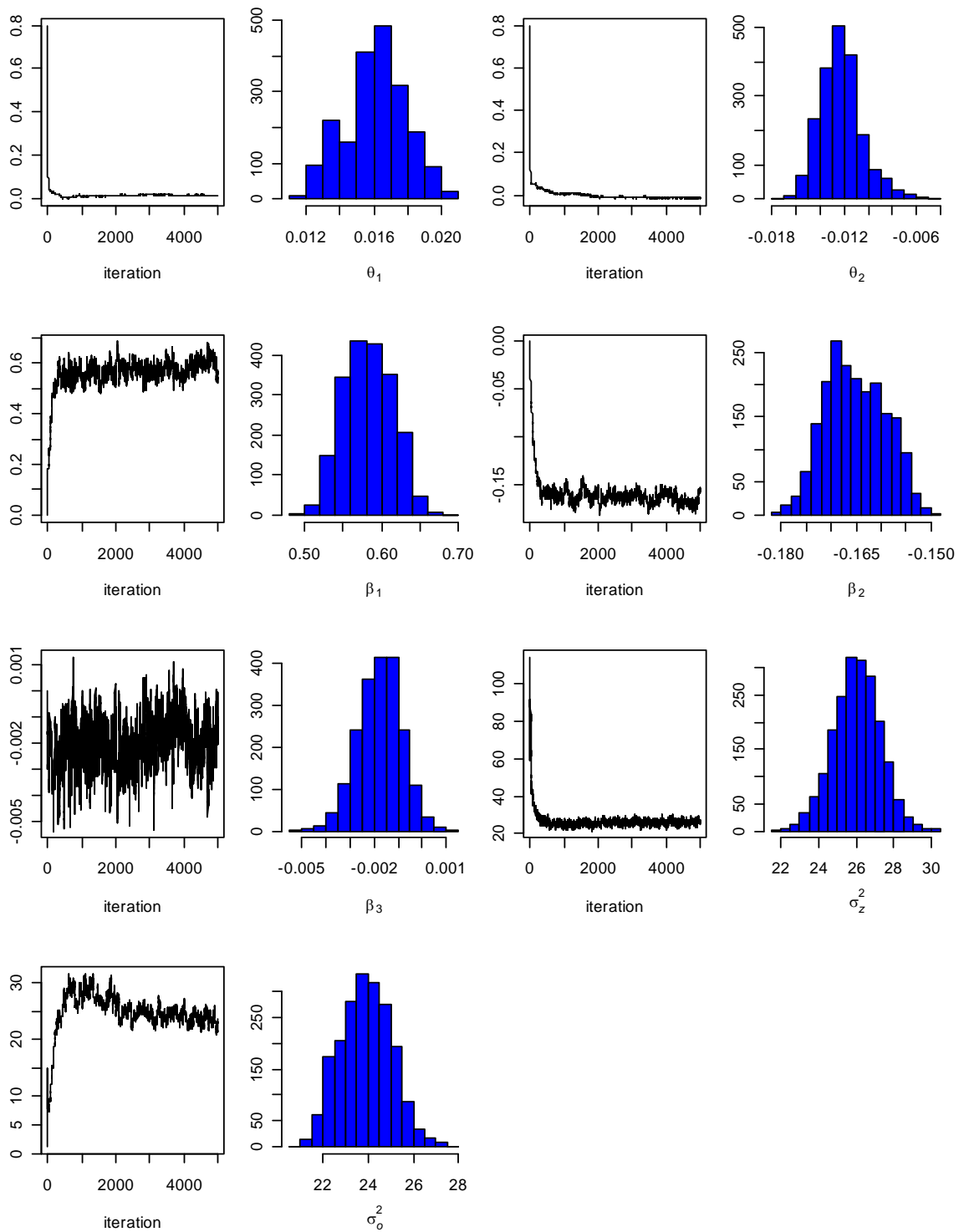
Parameter	<i>Model</i> <sub>3</sub>		<i>Model</i> <sub>4</sub>	
	Mean	s.d.	Mean	s.d.
$\phi_r$	(1.55, -0.34, 1)	(1.02, 0.47, 0.23)	(2.53, -0.34, 1)	(1.16, 0.44, 0.22)
$\phi_c$	(0.66, -0.4, 1)	(1.1, 0.42, 0.24)	(0.35, 0.12, 1)	(0.34, 0.53, 0.25)
$(\sigma_r^2, \sigma_{r0}^2)$	(0.11, 0.07)	(0.35, 0.09)	(0.64, 0.07)	(1.34, 0.08)
$(\sigma_c^2, \sigma_{c0}^2)$	(0.26, 0.08)	(1.22, 0.12)	(1.66, 0.08)	(3.06, 0.13)
$\sigma_b^2$	3	4.35	2	1.04

**Figure 4.8** Estimated surfaces of the autoregressive process  $\rho(\cdot)$  for models *Model*<sub>2</sub> and *Model*<sub>3</sub>.

The parameters' MCMC plots for model  $Model_2$  on the original data scale are presented in Figure 4.9. It appears that we have good convergence of the Gibbs sampler for all model parameters and the histograms of the posterior for the most of the parameters are concentrated in small intervals, except for the variance parameters,  $\sigma_z^2$  and  $\sigma_o^2$ , which have a wider spread. We have observed similar MCMC results for the other three models as well (not shown here).

In Chapter 3 we found that inference about the variance parameter can be very sensitive to the choice of the prior specification. Based on this, we performed additional runs of the models using different prior specifications. However, the results were not very sensitive to these changes in priors.





**Figure 4.9** Sequences of 5000 iterations from the posterior distributions of the model  $Model_2$  parameters and the histograms of the last 2000 iterations.

#### 4.4.2 Model performance

We characterize the predictive model performance by using root mean squared predictive error (RMSPE) as in Chapter 3 and we characterize overall model performance using average RMSE. The difference between RMSPE and RMSE is that in the first case we use one-step ahead model prediction values (2.22) and in the second case we use the model estimated concentration levels. These measures answer two different questions namely: how good is the forecast and how well does the model fit the data. For estimating the transport effect again we use the ratio of the variances of the transport term to the total variance. All figures reported in this section are based on the model *Model*<sub>2</sub>.

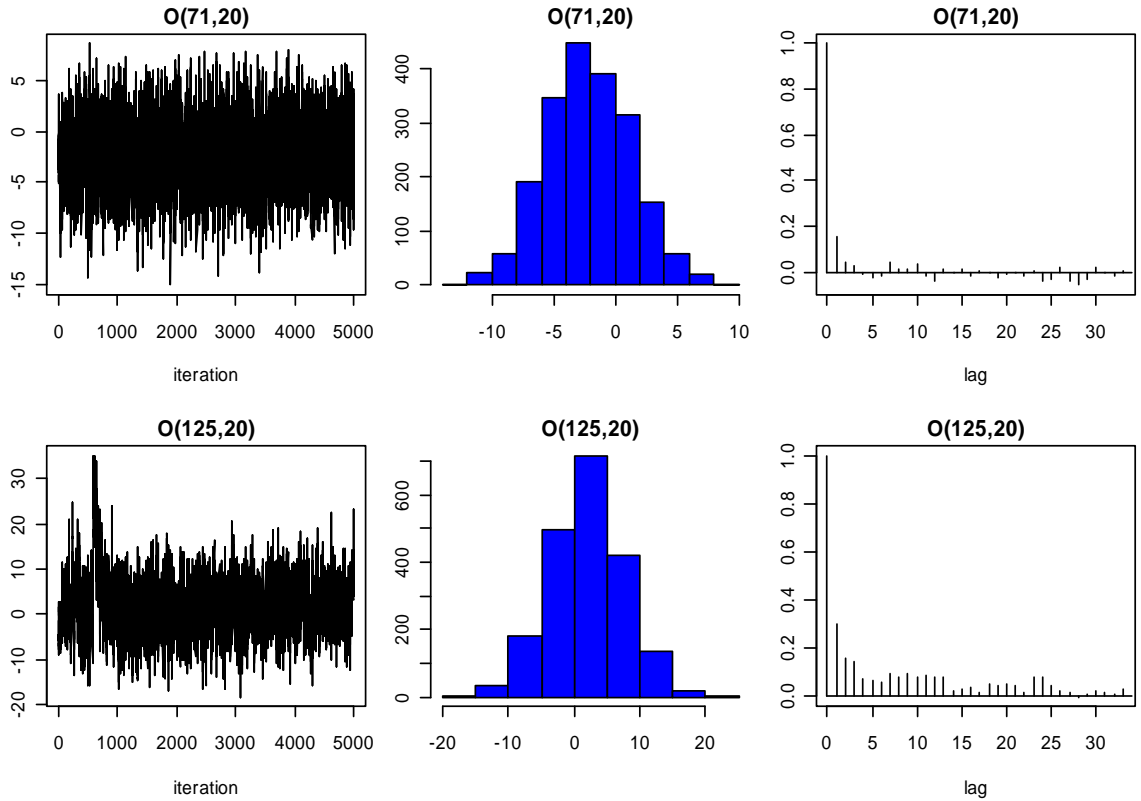
Table 4.3 reports the models overall performance and relative importance of the transport effect. We note that the average RMSEs stay at a very similar levels for all models. However, RMSPEs for models *Model*<sub>2</sub> and *Model*<sub>3</sub> that allow the time autoregressive processes to vary spatially are smaller than for models *Model*<sub>1</sub> and *Model*<sub>4</sub> as well as that transport effect for models *Model*<sub>2</sub> and *Model*<sub>3</sub> is larger than for the models *Model*<sub>1</sub> and *Model*<sub>4</sub>.

**Table 4.3** Summary of the overall models performance and the relative importance of the transport effect.

Model	RMSE [ppb]	RMSPE [ppb]	Tr. [%]
<i>Model</i> <sub>1</sub>	3.35	6.42	1%
<i>Model</i> <sub>2</sub>	3.54	6.01	4.2%
<i>Model</i> <sub>3</sub>	3.43	6.19	2.5%
<i>Model</i> <sub>4</sub>	3.34	6.44	0.4%

Next, we continue investigating the model performance on different scales. At the local scale, time point and grid cell level, our observations (Figure 4.10) indicate that higher uncertainty is associated with grid cells where no data are present. These results

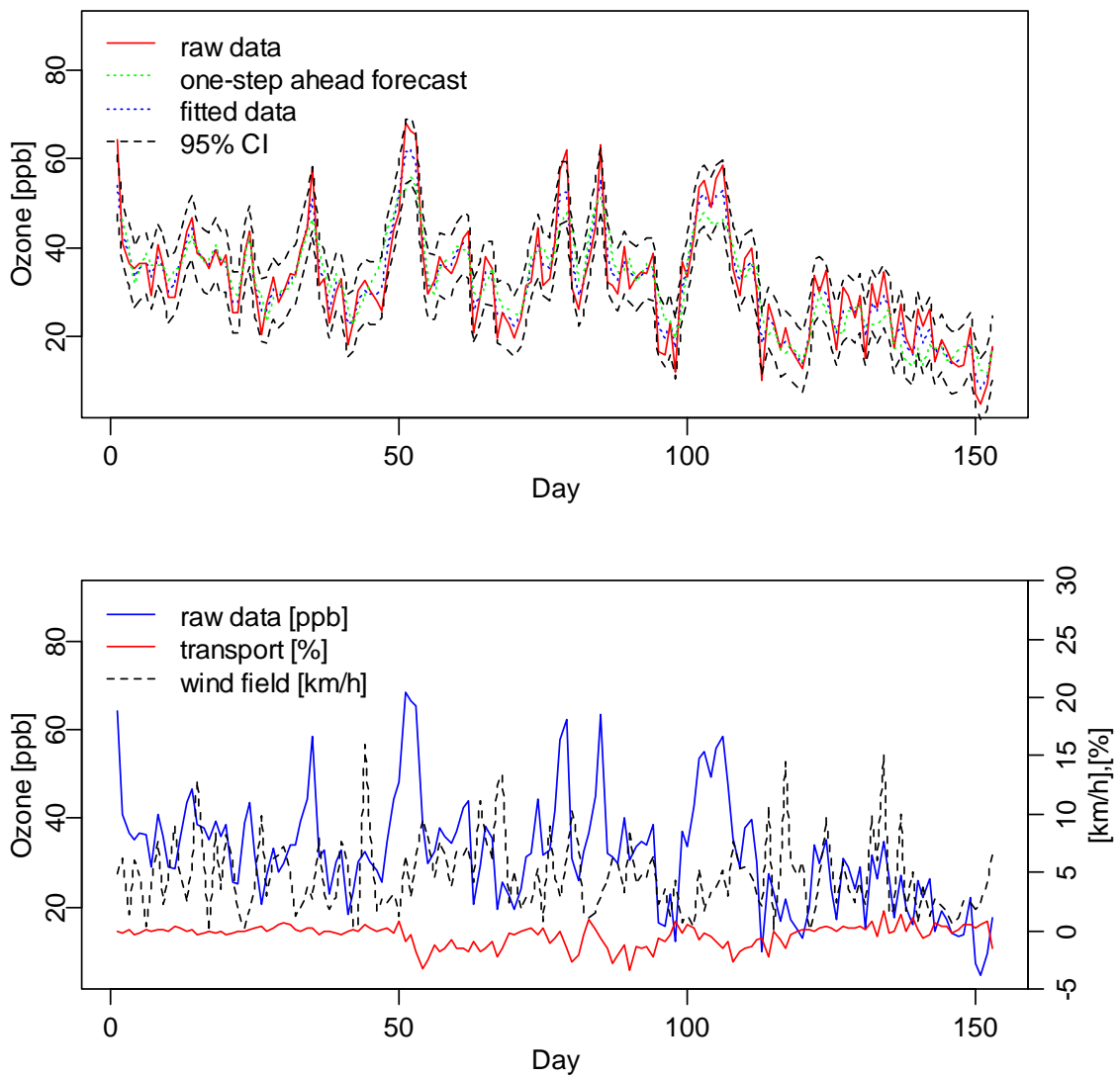
agree with our previous findings from Chapter 3. However, when we compare the MCMC results in Figure 3.4 and Figure 4.10 (left and right panels) it seems that autocorrelations among the Gibbs samples are smaller in the real data case. Although not shown, this behavior was observed throughout most of the prediction domain.



**Figure 4.10** Left: sequences of 5000 iterations from the posterior distributions of the randomly selected grid cells without observations  $O(125,20)$  and with observations  $O(71,20)$ ; Center: histogram of the last 2000 iterations; Right: autocorrelation function.

Figure 4.11 examines the model performance and the transport effect throughout the time period of study at a randomly selected location where data are present most of the time. That is, we investigate the model performance on domain-wide time scale and on the local scale in space. The top panel of the figure shows the observed ozone time series together with the estimated and one-step ahead predictions. We find the model in excellent agreement with the observed data and very good agreement between the pre-

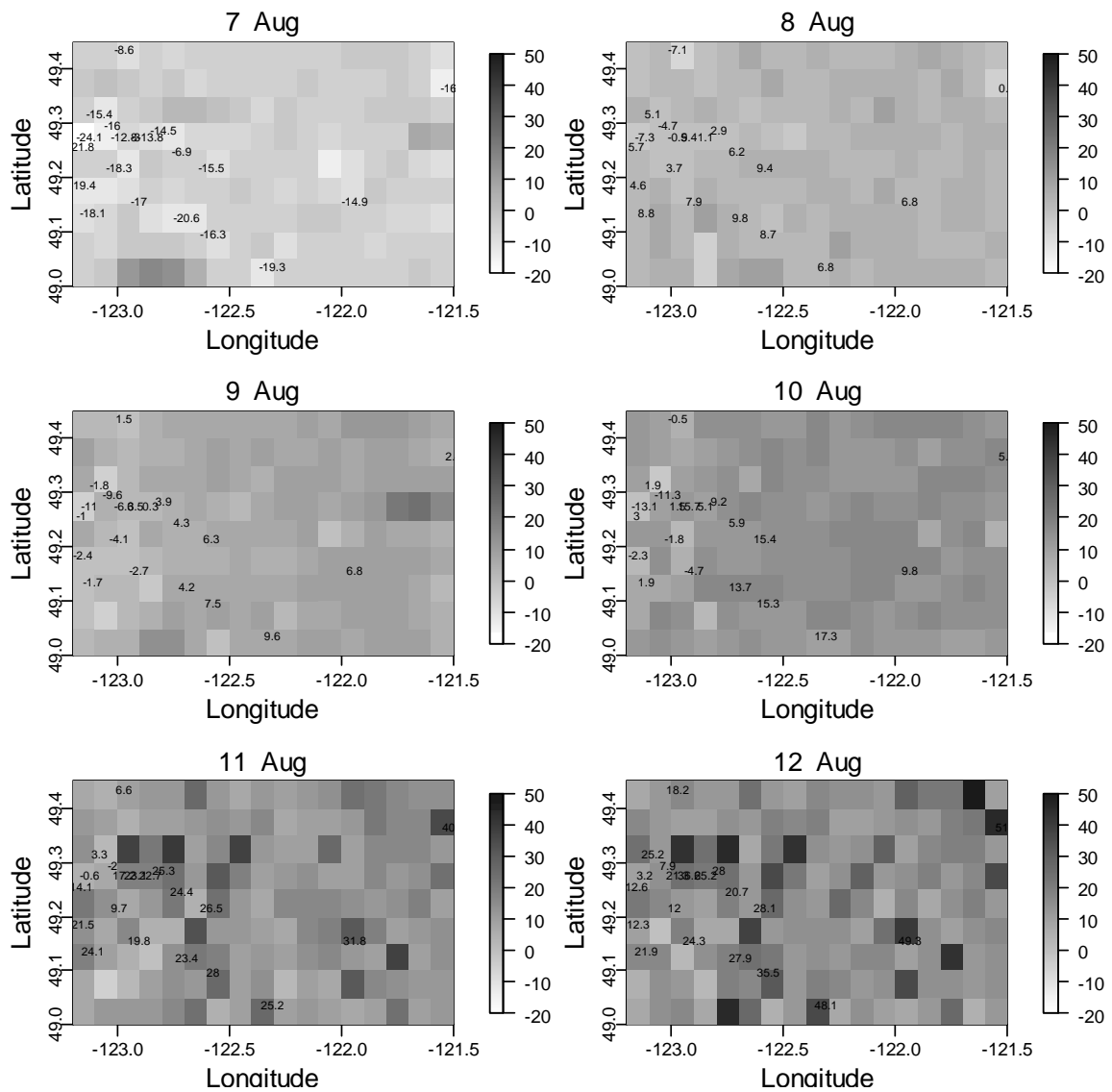
dicted and observed values. In both cases the observed time series is within the 95% pointwise credibility interval (CI). In the bottom panel of the figure we present the time series of the raw data, wind field and transport effect. As supported by our simulation analysis, the impact of the wind field on the ozone concentration levels can be positive, negative or zero. In addition, we found the difference in time evolution between grid cells with and without observation similar to the results in Chapter 3. Since these results are



**Figure 4.11** Time series for cells (75) with observations (station 30).

similar, we do not discuss them at length here. However, it should be noted that direct comparison between the results from the simulation study and real case results are not fully correct since the two lattices are quite different in size, 153 and 40 cells respectively. The reason I have used a smaller grid in the simulation study was purely practical. A real case model run takes about 5 to 7 hours and the Gibbs' sampler output is about 3.5 GB. In the simulation study I have to perform 32 model runs and to make the model development process feasible I chose to use a smaller grid.

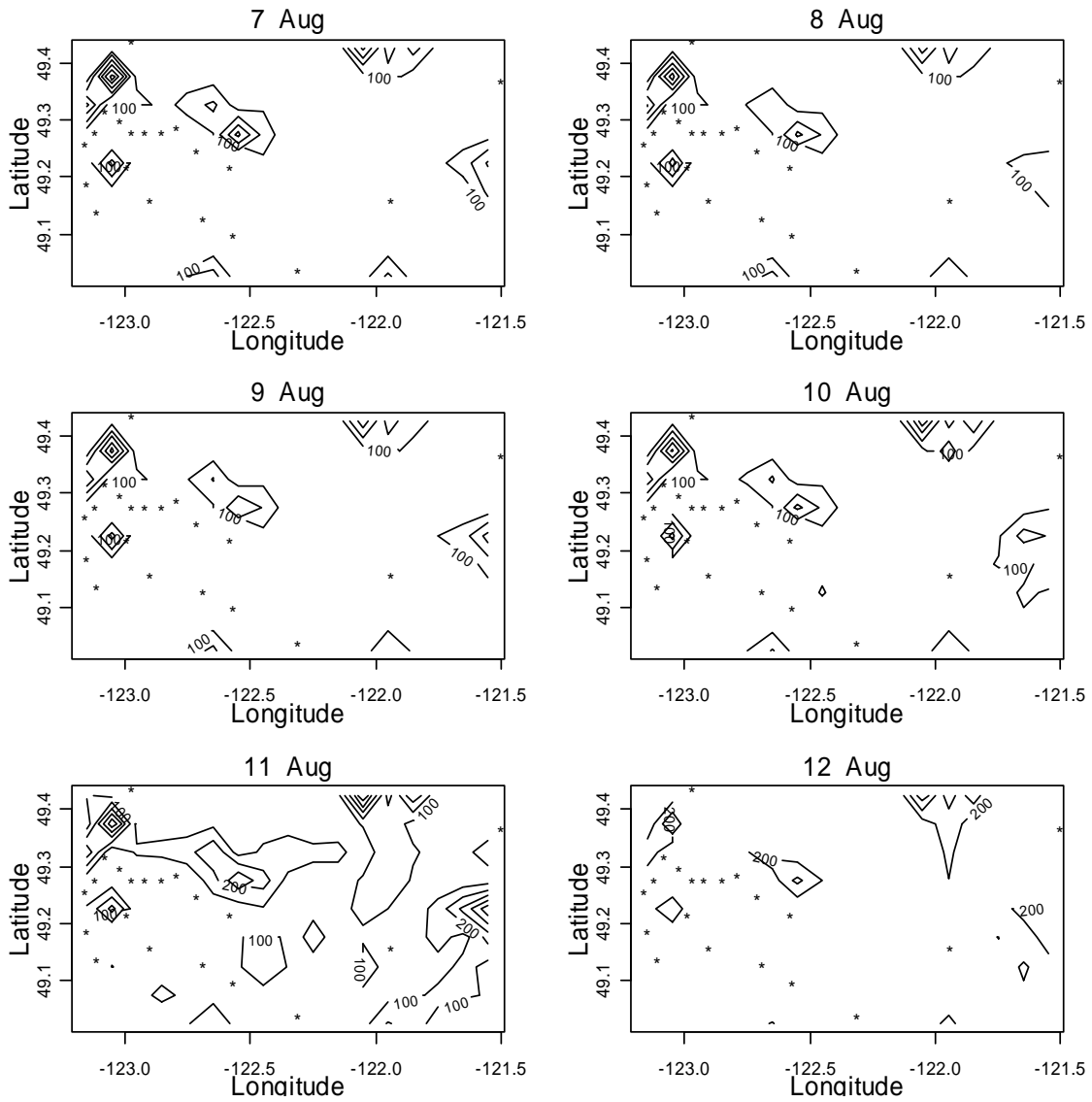
To investigate model performance on the domain scale both in time and space we used a sequence of model based spatial maps. Given the large number of time points over which the model was applied, we restrict our consideration to the period of 6 consecutive days, from 7 August to 12 August. This period comprises the most significant area-wide ozone episode observed in the region for the entire period of study (see Figure 4.4). Figure 4.12 displays 6 interpolated surfaces of the true process based on the model *Model*<sub>2</sub> as residuals. The model has correctly reproduced the observed space-time dynamic evolution and the model estimates were in close agreement with the observed values, superimposed on the plots. The RMSEs and RMSPEs in units of ppb for the period are 5.15, 5.02, 3.34, 6.33, 4.33, 6.81 and 9.04, 7.78, 5.65, 8.88, 9.12, 13.17 respectively. The average RMSE and RMSPE for the whole period are reported in Table 4.3. It should also be noted that we obtained similar results (not shown) from the other models. However, not surprisingly the spatial maps obtained with models *Model*<sub>1</sub> and *Model*<sub>4</sub> are smoother than the interpolated surfaces obtained with the models *Model*<sub>2</sub> and *Model*<sub>3</sub>. The latter models allow the time autoregressive process to vary spatially which would account for more realistic representation of the spatial variability.



**Figure 4.12** Model-estimated mean field of the true pollutant concentration as residuals. The observed values are superimposed on the plots.

Figure 4.13 displays corresponding standard deviation contour plots. The standard errors are smaller near the location of the observed data and basically trace the observed locations. These plots are not suitable for characterizing the domain variability for rea-

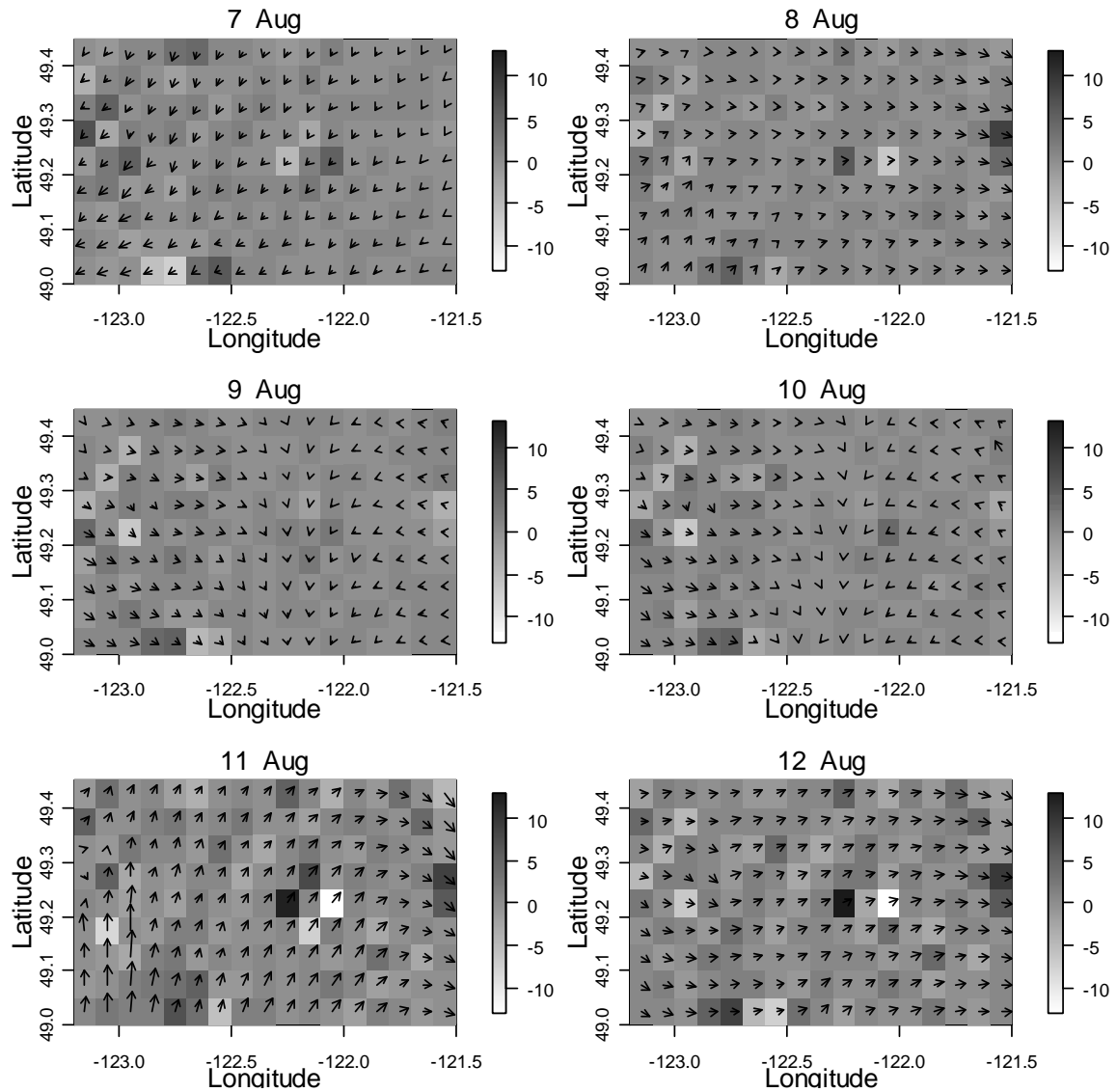
sons described in Chapter 3. Large-scale spatial variability is captured by the EOFs, and could be very different from those presented in Figure 4.13.



**Figure 4.13** Contour plots of pollutant variance field. \* - denotes the monitoring sites locations.

Figure 4.14 shows model-based interpolated transport effect maps. The domain mean transport effect is small and positive most of the time except day 3 when the effect is again small but negative. In the beginning of the ozone episode we have light easterly

winds with an average speed of about 3 km/h, and the smallest temperature for the ozone episode being analyzed. The ozone levels continue to increase as the temperature increases (see Figure 4.12) but the transport effect remains relatively constant.

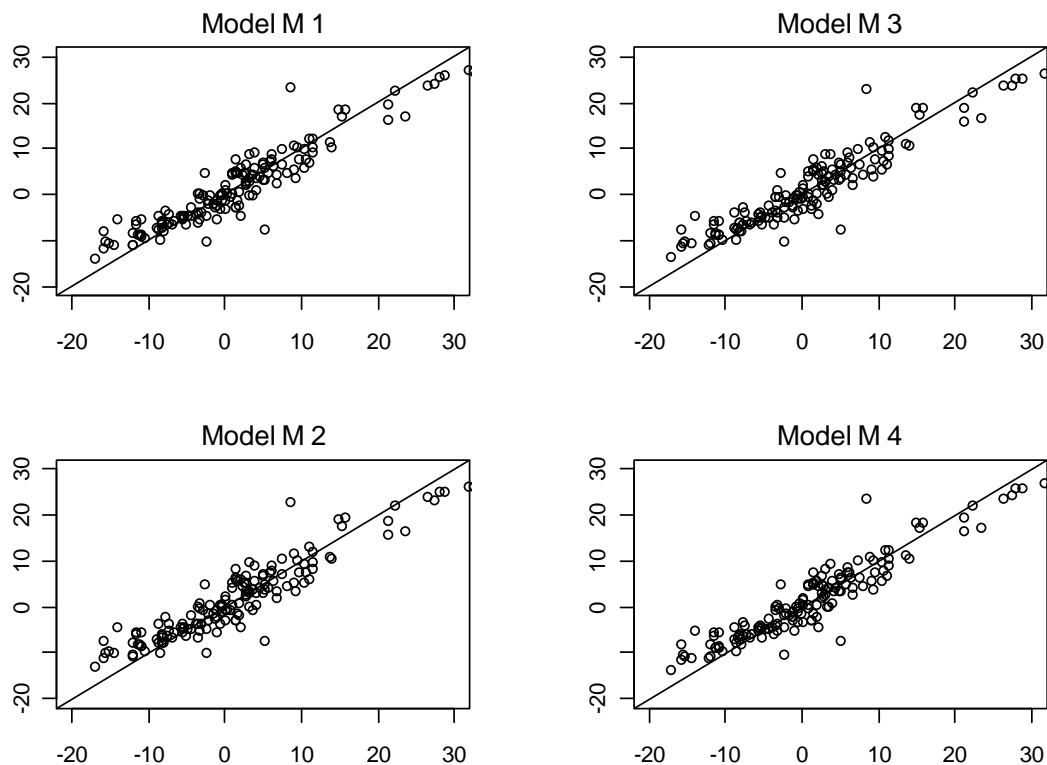


**Figure 4.14** Model estimated mean transport contributions map. The observed wind field is superimposed on the plots.



At the end of the episode the winds are light again: average wind speed about 2 km/h, but this time the winds are westerly. One further remark is that the maximum domain average temperature (29.05 °C) occurs on day 5 when the average wind speed is the highest for the episode, about 6 km/h, but the maximum ozone levels were observed on day 6 when we have light winds and still high temperatures (28.8 °C). This result suggests that the highest ozone levels are not generally associated with the higher wind speeds. This is in agreement with previous publications: Gardner and Dorling, 2000.

Figure 4.15 shows scatterplots of modeled and observed concentrations during the period of study at grid cell 75, obtained from various models. It can be seen that we have similar association between observed and modeled values for all models.



**Figure 4.15** Scatterplot of observed (x-axis) and modeled (y-axis) concentrations for models  $Model_1$ ,  $Model_2$ ,  $Model_3$ , and  $Model_4$  on the scale of residuals.

### 4.4.3 Model comparison

Our model provides an opportunity for partial comparison with the results obtained through simpler time series models that were employed in many previous ozone studies (Gardner and Dorling, 2000; McKendry, 2002; Robeson and Steyn, 1989, 1990; Pryor et. al, 1994; Pryor and Steyn, 1995). As a benchmark we used the two best models employed by Robson and Steyn (1989) namely: Temperature Persistent (TP) model and an ARIMA model. The TP model is a linear model that relates the ozone daily maximum value to the daily maximum air temperature ( $T$ ) and the previous day's daily maximum ozone concentration. That is:

$$O_t = b_0 + b_1 T_t + b_2 O_{t-1} + e_t \quad (4.2)$$

where  $e_t$  is the measurement error or unexplained residual term. This model can be seen as a special univariate case of the model (2.16) that takes into account partly the local production term (the regression part of (4.2)) and non-wind transport (autoregressive term in (4.2), persistence).

As in Robson and Steyn (1989) we analyze data from two monitoring stations, namely: station 09 (T09 in Robeson and Steyn, 1989) at Rocky Point Park and station 12 (T12) at Chilliwack Airport (see Fig. 4.2). These two stations crudely represent two distinct atmospheric situations: an urban (station T09) and a rural (station T12). Model performance statistics are summarized in Table 4.4 for comparative purposes. The comparison between TP and ARIMA model performance confirm the results in Robson and Steyn (1989). However, both models perform worse than all models introduced in Chapter 3. The models *Model*<sub>1</sub>, *Model*<sub>2</sub>, *Model*<sub>3</sub>, and *Model*<sub>4</sub> outperform the ARIMA and TP model by factor of 2-3. The average model RMSE of the spatiotemporal models is 3.7 ppb and the average RMSE of the univariate time series models is 8.6 ppb.

**Table 4.4** Comparison of model performance [ppb].

Station T9	Model*					
	(1)	(2)	(3)	(4)	(5)	(6)
Observed mean			-29.41-			
Observed standard deviation			-10.8-			
Predicted mean	29.6	29.9	29.7	28.9	29.5	29.46
Predicted standard deviation	13.3	7.0	7.06	8.08	9.8	7.93
RMSE	3.71	4.14	3.8	3.0	8.85	7.56

Station T12	Model					
	(1)	(2)	(3)	(4)	(5)	(6)
Observed mean			-33.96-			
Observed standard deviation			-13.4-			
Predicted mean	33.3	33.5	33.52	33.05	34.0	34.01
Predicted standard deviation	9.9	10.17	10.31	11.03	13.0	10.93
RMSE	4.08	4.36	4.06	3.22	10.25	7.55

\*(1) *Model*<sub>1</sub>, (2) *Model*<sub>2</sub>, (3) *Model*<sub>3</sub>, (4) *Model*<sub>4</sub>, (5) ARIMA(1,1,0), (6) Temperature and persistence based model

## 4.5 Discussion

In this chapter we have applied several space-time hierarchical models specified in the Chapter 3 to the LFV dataset. The hourly data, both meteorological and ozone were aggregated to daily rolling 8-hour averaged values. The meteorological data were interpolated using ordinary kriging over the entire domain of study for each particular day. However, only the daily maximum 8-hour average ozone values and corresponding meteorological fields were used in the analysis. Using 8-hour average values instead of hourly values simplifies (linearizes) the problem since we do not need to model small scale time signals (such as ozone diurnal cycle) present in the data.

Based on EDA we have specified the large-scale mean process. We modeled the mean space and time trend as an additive deterministic model. The time trend component was identified as a linear model and the large-scale spatial component we specified to be proportional to the first EOF mode of wind prefiltered ozone data. For this particular case we found the first eigenfunction to be the most appropriate choice to model a complex large-scale spatial pattern.

The models were able to capture a number of key ozone features such as: the strong dependence of ozone levels on temperature and persistence; the highest ozone levels are not generally associated with the strong winds; the wind's ability to transport pollutants. Additionally, we have investigated the models performance and ozone transport effect on different scales. The results show that the different models have similar performance and that the estimated transport effect could vary considerably. As we found in Chapter 3, the transport effect strongly depends on spatial data availability as well. For example, increasing the data availability from 20% to 80% could lead to an increase of 4 times of the estimated transport effect (Table 3.12). Hence, in this particular case, LFV dataset, where the data availability is only 11% the estimated overall transport effect (1%-4%) would be very erroneous.

Another characteristic of the models we used in this chapter is that we can attach uncertainty to all of these predictions and estimates. This uncertainty can be, derived directly from the model posterior distributions. As in Chapter 3 the uncertainty associated with the ozone level of a grid cell strongly depends on whether the data are presented or not present within and around that particular grid cell (Fig. 4.13). The uncertainty could decrease significantly if multiple monitoring sites are presented within a grid cell.

This chapter has been based on data from 2004. In order to broaden the analysis, we applied the model to data from 2005 and 2006. Model results and some summaries of the observed ozone and meteorological fields for years 2005 and 2006 are presented in Appendix B. We draw the same general conclusions we have made for 2004 as well as for 2005 and 2006. Additionally, we should point out that the model's performance remains relatively constant through the years as well as the most of the parameter estimates but there are some parameters that can vary significantly. The interannual variability of some of the model parameters is a result of the observed changes in the meteorological fields and ozone concentration levels. Investigation of the interannual variability (ozone trend) is beyond the scope of this study.

## **CHAPTER 5**

### **CONCLUSIONS AND FUTURE RESEARCH**

In this statistical study we have highlighted the importance of simultaneously considering different aspects of an air pollution problem as well as taking into account the physical bases that govern the processes of interest. We have developed a physically motivated stochastic model that decomposes the ground-level pollutant concentration field in three components, namely: transport, local production, and large-scale mean trend mostly dominated by the emission rates. The model is novel in the field of environmental spatial statistics and gives a different perspective on modeling in air pollution problem.

The transport component is modeled based on the advection equation. It was shown that different approximation schemes lead to different spatial autoregressive models. In our analysis we have used a lagged nearest-neighbor structure to model spatial dependency caused by the wind field. This is one of many possible parameterizations that could be investigated including simultaneous or staggered schemes. In Chapter 2 we showed that in modeling daily maximum 8-hour ozone data the simultaneous scheme is more appropriate. We have ignored the downward transport effect since: our dataset does not provide vertical wind fields and second intrusions of stratospheric ozone that reach the surface are rare events, though it is possible that surface concentrations can be affected by mixing air between the surface layer and the free troposphere above.

It is important to emphasize two factors when modeling the transport of an air pollutant: wind field and the pollutant gradients should be considered and they should be considered concurrently. Excluding small-scale phenomena such as diffusion, means it is not possible to have transport if there are no gradients in the pollutant field and/or there is

no wind. This fact is a consequence of the advection equation and has been employed by the proposed models.

The local production component is implemented as a vector regression model with fixed in time and space coefficients. Adopting such an approach allows us to ignore contributions from the neighboring sites and takes into account only specific local meteorological conditions. This puts physical/chemical meaning in our regression coefficients, which is a desirable feature of a local production model since physics and chemistry do not depend on location. Therefore, the local production spatial structure is not modeled directly, but is inherent in the covariates field.

The meteorological fields used in this study were obtained through a kriging method that interpolated meteorological data over the grid. That is, the meteorological variables, which are required for the model as input, are not modeled dynamically. This together with the poor representativeness of the meteorological data could lead to poorer model performance. In forecasting mode the model requires dynamic meteorological fields as starting input in order to predict ozone concentrations. Alternatively, the meteorological fields can be obtained from weather forecast data.

The large-scale mean process is specified based on the results obtained from EDA. In EDA we have used wind prefiltered ozone data to determine the first EOF mode that is used as large-scale spatial trend mostly dominated by the emissions rates. This is a new approach that allows us to avoid using emissions inventory data. However, the approach is only applicable if rich datasets are available. Further study is needed to improve the methods for estimation of the emission rates maps as well as to investigate the dependence of ozone transport on the details of wind fields.

The proposed models were able to capture a number of key features of ozone behavior and show excellent agreement with the observed data. The models outperformed the considerably simpler univariate time series models by a factor of 2 to 3. Additionally, the Bayesian framework allowed us to easily incorporate the scientific knowledge and to

estimate grid-cell average values, conditioned on point-referenced observed data. recently, Berliner (2003) emphasized that stochastic and deterministic models are not separate items, but should be considered ends of a continuum. In this work we have demonstrated this by providing a stochastic model that captures deterministic elements of the problem. More specific discussions and conclusions are provided in the discussion subsections of Chapter 3 and 4.

Finally, more or less any aspect of this approach can be further developed. This includes: further exploring ground-level ozone modeling on point-level spatial resolution that takes into account pollutant transport; considering simultaneous nearest neighbor structure or other approximation schemes for modeling the transport component; improving the methods for estimation of the emission rates; coupling the proposed model with the output of a Numerical Weather Prediction (NWP) system; improving the way relative importance of the transport effect is calculated.



## REFERENCES

- Ainslie B. and Steyn D.G., 2007. Spatio-Temporal Trends in Episodic Ozone Pollution in the Lower Fraser Valley, B.C. in Relation to Meso-scale Atmospheric Circulation Patterns and Emissions. *In Press. Journal of Applied Meteorology*. (October, 2006).
- Banerjee S., Carlin B.P., and Gelfand A.E., 2004. Hierarchical Modeling and Analysis for Spatial Data. *Chapman and Hall/CRC, Boca Raton, FL*.
- Bennett R.J., 1979, Spatial time series: Analysis-forecasting-control, *Pion, London, 674p*.
- Berliner L.M., 2003. Physical-statistical modeling in geophysics, *Journal of Geophysical Research*, vol. 108, no. d24, 8776, doi:10.1029/2002jd002865.
- Besag J., 1974, Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B*, 36:192-225.
- Box G. and Jenkins G., 1976. Time series analysis forecasting and control, *Holden-Day, Inc.*
- Breiman L., Friedman J.H., Olshen R.A. and Stone C.J., 1984. Classification and Regression Trees. *Wadsworth and Brooks/Cole, Monterey*.
- Bretthorst L., 1988. Bayesian Spectral Analysis and Estimation. *New York, Springer*.
- Burrows W.R., Benjamin M., Beauchamp S., Lord E.R., McCollor D. and Thomson B., 1995. CART decision tree statistical analysis and prediction of summer season maximum surface ozone for the Vancouver, Montreal, and Atlantic Regions of Canada. *Journal of Applied Meteorology* 34, 1848-1862.
- Byun, D., Young, J., Gipson, J., Godowitch, J., Binkowski, F., Roselle, S., Benjey, B., Pleim, J., Ching, J., Novak, J., Coats, C., Odman, T., Hanna, A., Alapaty, A., Mathur, R., McHenry, J., Sankar, U., Fine, S., Xiu, A. and Jang, C., 1998, *Description of the Models-3 Community Multiscale Air Quality (CMAQ) model*. Proceedings of the American Meteorological Society 78th Annual Meeting, January 11–16, Phoenix, AZ.
- Carroll R.J., Chen R., George E.I., Li T.H., Newton H.J., Schmiediche H., Wang N., 1997. Ozone exposure and population density in Harris county, Texas. *Journal of the American Statistical Association* 92: 392–415.

Carter C.K. and Kohn R., 1994. On gibbs sampling for state space models. *Biometrika*, 81(3).

Cobourn W.G. and Hubbard M.C., 1999. An enhanced ozone forecasting model using air mass trajectory analysis, *Atmospheric Environment*, Vol. 33, pp. 4663–4674.

Cox W. and Chu S., 1996. Assessment of interannual ozone variation in urban areas from a climatological perspective. *Atmospheric Environment*, Vol. 30, p. 2615-2625.

Cressie N. and Huang H-C., 1999. Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association* 94: 1330–1340.

Cressie N., 1993. Statistics for Spatial Data, *John Wiley & Sons, INC*.

Christakos G., 1991. On certain classes of spatiotemporal random fields with applications to space - time data processing. *IEEE Transactions on Systems, Man, and Cybernetics*, 21, 861 - 875.

Fuentes M., Chen L., and Davis J.M., 2008. A class of nonseparable and nonstationary spatial temporal covariance functions. *Environmetrics* 2008; 19: 487–507

Fuentes M., Guttorp P., and Sampson P., 2006. Using transforms to analyze space-time processes. *Chapter in Finkelsted et al.(eds), pp. 77-150.CRC/Chapman Hall*.

Gardner M.W., Dorling S.R., 2000. Statistical surface ozone models: an improved methodology to account for non-linear behavior, *Atmospheric Environment*, Vol. 34, p. 21-34.

Gelman A. and Rubin D.B., 1992. Inference from iterative simulation using multiple sequences, *Statistical Science*, 7, 457-511.

Gilks W.R., Richardson S., and Spiegelhalter D.J., 1996. Markov Chain Monte Carlo in Practice. *Chapman and Hall/CRC, Boca Raton, FL*.

Gregory P.C., 2005. Bayesian Logical Data Analysis for Physical Sciences: A Comparative Approach with Mathematica Support. *Cambridge University Press, 589 pp*.

Gneiting T., 2002. Nonseparable covariance function for space-time data. *Journal of the American Statistical Association* 97: 590–600.

Gumbel E.J., 1958. Statistics and Extremes, *Columbia University Press, New York*.

Guttorp P., Meiring W., and Sampson P.D., 1994. A space–time analysis of ground-level ozone data. *Environmetrics* 5: 241–254.

Government of British Columbia, Ministry of environment, 2005. Lower Fraser Valley Ambient Air Quality Report. [http://wlapwww.gov.bc.ca/sry/p2/air\\_quality/index.html](http://wlapwww.gov.bc.ca/sry/p2/air_quality/index.html)

Hass H., 1991. Description of the EURAD chemistry transport module (CTM) version 2. In: Ebel, A., Neubauer, F.M., Speth, P. (Eds.), Report 83. Institute of Geophysics and Meteorology, University of Cologne, Cologne, Germany.

Holmes N.S. and Morawska L., 2006. A review of dispersion modeling and its application to the dispersion of particles: An overview of different dispersion models available. *Atmospheric Environment*, Volume 40, Issue 30, p. 5902-5928.

Hsieh W.W., 2004. Nonlinear multivariate and time series analysis by neural network methods. *Reviews of Geophysics*, 42, RG1003, doi:10.1029/2002RG000112

Huang H-C. and Hsu N-J., 2004, Modeling transport effects on ground-level ozone using a non-stationary space–time model, *Environmetrics*, 15: 251–268 (DOI: 10.1002/env.639).

Huerta G., Sanso B., and Stroud J.R., 2004. A spatiotemporal model for Mexico City ozone levels, *Appl. Statist.*, 53, Part 2, pp. 231–248.

Jimenez P., Baldasano, J. M., Dabdub D., 2003. Comparison of photochemical mechanisms for air quality modeling. *Atmospheric Environment*. Vol.37: p.4179-4194.

Joe H., Steyn D.G. and Susko E., 1996. Analysis of trends in tropospheric ozone in the Lower Fraser Valley, British Columbia. *Atmospheric Environment*, Vol. 30, No. 20, p. 3413-3423.

Kalnay, E. (2003), Atmospheric modeling, data assimilation and predictability, Cambridge: Cambridge University Press

Karl T.R., Koscielny A.J., and Diaz H.F., 1982. Potential Errors in the Application of Principal Component (Eigenvector) Analysis to Geophysical Data, *Journal of Applied Meteorology* Volume 21, Issue 8.

Le N.D. and Zidek J.V., 2006. Statistical analysis of environmental space-time processes. New York: Springer.

- Le N.D., Li S., Zidek J. V., 2001. Spatial Prediction and Temporal Backcasting for Environmental Fields Having Monotone Data Patterns, *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, Vol. 29, No. 4., pp. 529-554.
- Lindley D.V. and Smith A.F., 1972. Bayes estimates for the linear model. *Journal of the Royal Statistical Society B*, 34A.
- Lorenz E.N., 1963. Deterministic nonperiodic flow, *Journal of Atmospheric Sciences*, Vol. 20.
- Lorenz E. N., 1956. Empirical orthogonal functions and statistical weather prediction. *Sci. Rep. No. 1, Statistical Forecasting Project, M.I.T., Cambridge, MA*, 48 pp.
- McKendry I.G., 2002. Evaluation of Artificial Neural Networks for fine particulate pollution (PM<sub>10</sub> and PM<sub>2.5</sub>) forecasting, *J Air & Waste Manage Assoc* 52, 1096–1101.
- McMillan N., Bortnick S.M., Irwin M.E., Berliner L.M., 2005. A hierarchical Bayesian model to estimate and forecast ozone through space and time. *Atmospheric Environment* 39, pp. 1373–1382.
- McPhaden M. J. and Hayes S.P., 1991. On the Variability of Winds, Sea Surface Temperature, and Surface Layer Heat Content in the Western Equatorial Pacific. *Journal of Geophysical Research*, 96, supplement, 3331-3342.
- NRC (National Research Council), 1991, Rethinking the Ozone Problem in Urban and Regional Air Pollution. *National Academy Press, Washington*.
- Odman M.T., Ingram C.L., 1996. Multiscale Air Quality Simulation Platform (MAQSIP). Source Code Documentation and Validation. *MCNC Report No. ENV-96TR002-v 1.0. MCNC, Research Triangle Park, North Carolina*.
- Понтрягин Л.С., Андронов А.А., Витт А.А., 1933. О статистическом рассмотрении динамических систем. — *Журнал Экспериментальной и Теоретической Физики*, т. 3, вып. 3, с. 165—180. Андронов А. А. *Собр. трудов*, с. 142—160.
- Pryor S.C., McKendry I.G., and Steyn D.G., 1994. Synoptic-scale meteorological variability and surface ozone concentrations in Vancouver, BC. *Journal of Applied Meteorology*, Vol. 34, No. 8, p. 1824-1833.
- Pryor S.C. and Steyn D.G., 1995. Hebdomadal and diurnal cycles in ozone time series from the Lower Fraser Valley, BC. *Atmospheric Environment*, Vol. 29, No. 9, p. 1007-1019.

- Randall D.A., 2006. Reynolds Averaging. [http://kiwi.atmos.colostate.edu/group/dave/pdf/Reynolds\\_Averaging.pdf](http://kiwi.atmos.colostate.edu/group/dave/pdf/Reynolds_Averaging.pdf)
- Rao S.T., Zurbenko I.G., 1994. Detecting and Tracking Changes in Ozone Air Quality. *J Air & Waste Manage Assoc* 44, 1089–1092.
- Riccio A., Barone G., Chianese E., and Giunta G., 2006, A hierarchical Bayesian approach to the spatio-temporal modeling of air quality data. *Atmospheric Environment, Vol. 40, 554–566.*
- Robeson, S.M. and Steyn D.G., 1989. A Conditional Probability Density Function for Ozone Air Quality Data," *Atmospheric Environment, 23 (3), p. 689-692.*
- Robeson S.M. and Steyn D.G., 1990. Evaluation of Statistical Forecast Models for Daily Maximum Ozone Concentrations. *Urban Atmosphere, 24b, p. 303-312.*
- Royle, J. A., Berliner, L.M., Wikle C. K., and Milliff R., 1998. A hierarchical spatial model for constructing wind fields from scatterometer data in the Labrador Sea. *Case Studies in Bayesian Statistics*, edited by C. Gatsonis et al., pp. 367– 382, Springer-Verlag, New York.
- Royle, J. A. and Berliner, L.M., 1999. A Hierarchical Approach to Multivariate Spatial Modeling and Prediction. *Journal of Agricultural, Biological, and Environmental Statistics*. Vol. 4, No. 1, pp. 29–56.
- Russell A. and Dennis R., 2000. NARSTO critical review of photochemical models and modeling. *Atmospheric Environment, 2000, Vol.34, p. 2283-2324.*
- Schabenberg O. and Gotway C.A., 2005. Statistical Methods for Spatial Data Analysis. *Chapman and Hall/CRC, Boca Raton, FL.*
- Sahu S.K., Gelfand A.E., and Holland D.M., 2007. High-Resolution Space–Time Ozone Modeling for Assessing Trends. *Journal of the American Statistical Association* December 2007, Vol. 102, No. 480
- Seinfeld J.H., Pandis S.N., 1998, Atmospheric chemistry and physics, *John Wiley & Sons, INC.*
- Shigeyuki O., Masa-aki S., Ichiro T., Morito M., Ken-ichi M., and Shin I., 2003. A Bayesian missing value estimation method for gene expression profile data. *Bioinformatics, 19(16):2088-2096.*
- Smith R.L. and Huang L., 1993. Modeling high threshold exceedances of urban ozone.

National Institute for Statistical Science Technical Report #6.

Simard Y. and Marcotte D., 1993. Assessing similarities and differences among maps: A study of temporal changes in distribution of northern shrimp (*pandalus borealis*) in the gulf of St. Lawrence, in A. Soares, ed., *Geostatistics Troia '92*, Vol. 2: Kluwer Academic Publ., Dordrecht, p. 865–874.

Simpson D., 1995. Biogenic emissions in Europe 2: Implications for ozone control strategies. *Journal of Geophysical Research* 100, 22891-22906.

Stroud J. R., Muller P., and Sanso B., 200. Dynamic models for spatiotemporal data. *Journal of Royal Statistical Society, Series B*, **63**, 673- 689.

Stull B.R., 1988. *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers.

Thompson M.L., Reynolds J., Cox L.H., Guttorp P. and Sampson P.D., 2001. A review of statistical methods for the meteorological adjustment of tropospheric ozone. *Atmospheric Environment* **35**, pp. 617–630.

Tobías A. and Scotto M.G., 2005. Prediction of extreme ozone levels in Barcelona, Spain, *Environmental Monitoring and Assessment* , Volume 100, Numbers 1-3.

Vingarzan R. and Taylor B., 2003. Trend analysis of ground level ozone in the greater Vancouver/Fraser Valley area of British Columbia, *Atmospheric Environment*, Vol. 37, p. 2159–2171.

von Storch H. and Zwiers F., 2000. *Statistical Analysis in Climate Research*, Cambridge University Press, 476 pp.

West M. and Harrison P. J., 1997. *Bayesian Forecasting and Dynamic Models*. Springer: New York, 2nd ed.

Wikle C.K. (1996), Spatio-temporal statistical models with application to atmospheric processes, Ph.D. dissertation, *Iowa State University, Ames, IA*.

Wikle C.K., Berliner M. and Cressie N., 1998, Hierarchical Bayesian space-time models. *J. Environ. Ecol. Statist.*, 5, 117–154.

Wikle C.K., Cressie N., 1999. A dimension reduction approach to space-time Kalman filtering. *Biometrika* 86, 815–829.

Whittle P., 1954. On stationary processes in the plan. *Biometrika*, 41:434-449

Yin D., Jiang W., Roth H., and Giroux E., 2004. Improvement of biogenic emissions estimation in the Canadian Lower Fraser Valley and its impact on particulate matter modeling results. *Atmospheric Environment*, Vol.38(4), p. 507-521

## APPENDIX A

### Full conditional distributions for Gibbs sampling

The Gibbs sampling is the simplest and the most easily implemented algorithm for doing MCMC computer simulations. To implement a Gibbs sampler we need to know the full conditional distributions of all unknowns. However, it turns out that we only need to know the kernel part of these distributions and do not need to know the normalizing constants which often does not have analytical expression. That is why the full conditional distributions derived here are only known up to a constant factor.

$\mathbf{O}_t$ 's full conditional distribution follows from (2.1) and (3.2):

$$\begin{aligned}
[\mathbf{O}_t | \cdot] &\propto [\mathbf{Z}_t | K, \mathbf{O}_t, \sigma_z^2][\mathbf{O}_t | M_t, \boldsymbol{\beta}, \boldsymbol{\rho}, H_{t-1}, H_{it}^B, \mathbf{O}_{t-1}, \mathbf{O}_{it}^B, \sigma_o^2][\mathbf{O}_{t+1} | M_t, \boldsymbol{\beta}, \boldsymbol{\rho}, H_t, H_{it}^B, \mathbf{O}_t, \mathbf{O}_{it}^B, \sigma_o^2] \\
&\propto \exp \left\{ -\frac{1}{2\sigma_z^2} (\mathbf{Z}_t - K\mathbf{O}_t)' (\mathbf{Z}_t - K\mathbf{O}_t) \right. \\
&\quad -\frac{1}{2\sigma_o^2} (\mathbf{O}_t - M_t\boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{t-1})\mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B)' \\
&\quad \quad \times (\mathbf{O}_t - M_t\boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{t-1})\mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B) \\
&\quad -\frac{1}{2\sigma_o^2} (\mathbf{O}_{t+1} - M_{t+1}\boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_t)\mathbf{O}_t - \sum_{i=1}^4 H_{it+1}^B \mathbf{O}_{it+1}^B)' \\
&\quad \quad \times (\mathbf{O}_{t+1} - M_{t+1}\boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_t)\mathbf{O}_t - \sum_{i=1}^4 H_{it+1}^B \mathbf{O}_{it+1}^B) \left. \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[ \mathbf{O}_t' \left( \frac{1}{\sigma_z^2} K'K + \frac{1}{\sigma_o^2} (I_N + \text{diag}(\boldsymbol{\rho}) + H_t)' (I_N + \text{diag}(\boldsymbol{\rho}) + H_t) \right) \mathbf{O}_t \right. \right. \\
&\quad -2 \left( \frac{1}{\sigma_z^2} K'\mathbf{Z}_t + \frac{1}{\sigma_o^2} (M_t\boldsymbol{\beta} + (\text{diag}(\boldsymbol{\rho}) + H_{t-1})\mathbf{O}_{t-1} + \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B) \right. \\
&\quad \quad \left. \left. + \frac{1}{\sigma_o^2} (\text{diag}(\boldsymbol{\rho}) + H_t)' (\mathbf{O}_{t+1} - M_{t+1}\boldsymbol{\beta} - \sum_{i=1}^4 H_{it+1}^B \mathbf{O}_{it+1}^B) \right)' \mathbf{O}_t \left. \right\}
\end{aligned}$$

$t = 2, \dots, T - 1$ . Thus

$$\begin{aligned}
[\mathbf{O}_t | \cdot] &\sim MVN(\Lambda_t \chi_t, \Lambda_t) \quad \text{where} \\
\Lambda_t^{-1} &= \left( \frac{1}{\sigma_z^2} K'K + \frac{1}{\sigma_o^2} I_N + \frac{1}{\sigma_o^2} (\text{diag}(\boldsymbol{\rho}) + H_t)' (\text{diag}(\boldsymbol{\rho}) + H_t) \right)
\end{aligned}$$

and



$$\begin{aligned}\chi_t = & \left( \frac{1}{\sigma_z^2} K' \mathbf{Z}_t + \frac{1}{\sigma_o^2} (M_t \boldsymbol{\beta} + (\text{diag}(\boldsymbol{\rho}) + H_{t-1}) \mathbf{O}_{t-1} + \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B) \right. \\ & \left. + \frac{1}{\sigma_o^2} (\text{diag}(\boldsymbol{\rho}) + H_t)' (\mathbf{O}_{t+1} - M_{t+1} \boldsymbol{\beta} - \sum_{i=1}^4 H_{it+1}^B \mathbf{O}_{it+1}^B) \right)\end{aligned}$$

We need to “initialize” the time series. The full conditional distribution of  $\mathbf{O}_1$  is given by

$$\begin{aligned}[\mathbf{O}_1 | \cdot] & \propto [\mathbf{Z}_1 | K, \mathbf{O}_1, \sigma_z^2] [\mathbf{O}_1 | M_1, \boldsymbol{\beta}, H_{i1}^B, \mathbf{O}_{i1}^B, \sigma_o^2] [\mathbf{O}_2 | M_2, \boldsymbol{\beta}, \boldsymbol{\rho}, H_1, H_{i2}^B, \mathbf{O}_1, \mathbf{O}_{i2}^B, \sigma_o^2] \\ & \propto \exp \left\{ -\frac{1}{2\sigma_z^2} (\mathbf{Z}_1 - K \mathbf{O}_1)' (\mathbf{Z}_1 - K \mathbf{O}_1) \right. \\ & \quad - \frac{1}{2\sigma_o^2} (\mathbf{O}_1 - M_1 \boldsymbol{\beta} - \sum_{i=1}^4 H_{i1}^B \mathbf{O}_{i1}^B)' \\ & \quad \quad \times (\mathbf{O}_1 - M_1 \boldsymbol{\beta} - \sum_{i=1}^4 H_{i1}^B \mathbf{O}_{i1}^B) \\ & \quad - \frac{1}{2\sigma_o^2} (\mathbf{O}_2 - M_2 \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_1) \mathbf{O}_1 - \sum_{i=1}^4 H_{i2}^B \mathbf{O}_{i2}^B)' \\ & \quad \quad \times (\mathbf{O}_2 - M_2 \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_1) \mathbf{O}_1 - \sum_{i=1}^4 H_{i2}^B \mathbf{O}_{i2}^B) \left. \right\} \\ & \propto \exp \left\{ -\frac{1}{2} \left[ \mathbf{O}_1' \left( \frac{1}{\sigma_z^2} K' K + \frac{1}{\sigma_o^2} (I_N + \text{diag}(\boldsymbol{\rho}) + H_1)' (I_N + \text{diag}(\boldsymbol{\rho}) + H_1) \right) \mathbf{O}_1 \right. \right. \\ & \quad - 2 \left( \frac{1}{\sigma_z^2} K' \mathbf{Z}_1 + \frac{1}{\sigma_o^2} (M_1 \boldsymbol{\beta} + \sum_{i=1}^4 H_{i1}^B \mathbf{O}_{i1}^B) \right. \\ & \quad \quad \left. \left. + \frac{1}{\sigma_o^2} (\text{diag}(\boldsymbol{\rho}) + H_1)' (\mathbf{O}_2 - M_2 \boldsymbol{\beta} - \sum_{i=1}^4 H_{i2}^B \mathbf{O}_{i2}^B) \right)' \mathbf{O}_1 \left. \right\}\end{aligned}$$

Thus

$$\begin{aligned}[\mathbf{O}_1 | \cdot] & \sim MVN(\Lambda_1 \chi_1, \Lambda_1) \quad \text{where} \\ \Lambda_1^{-1} & = \left( \frac{1}{\sigma_z^2} K' K + \frac{1}{\sigma_o^2} (I_N + \text{diag}(\boldsymbol{\rho}) + H_1)' (I_N + \text{diag}(\boldsymbol{\rho}) + H_1) \right)\end{aligned}$$

and

$$\chi_1 = \left( \frac{1}{\sigma_z^2} K' \mathbf{Z}_1 + \frac{1}{\sigma_o^2} (M_1 \boldsymbol{\beta} + \sum_{i=1}^4 H_{i1}^B \mathbf{O}_{i1}^B) + \frac{1}{\sigma_o^2} (\text{diag}(\boldsymbol{\rho}) + H_1)' (\mathbf{O}_2 - M_2 \boldsymbol{\beta} - \sum_{i=1}^4 H_{i2}^B \mathbf{O}_{i2}^B) \right)$$

Similar for  $t = T$

$$\begin{aligned}[\mathbf{O}_T | \cdot] & \propto [\mathbf{Z}_T | K, \mathbf{O}_T, \sigma_z^2] [\mathbf{O}_T | M_T, \boldsymbol{\beta}, \boldsymbol{\rho}, H_{iT}, H_{iT}^B, \mathbf{O}_{iT-1}, \mathbf{O}_{iT}^B, \sigma_o^2] \\ & \propto \exp \left\{ -\frac{1}{2\sigma_z^2} (\mathbf{Z}_T - K \mathbf{O}_T)' (\mathbf{Z}_T - K \mathbf{O}_T) \right. \\ & \quad - \frac{1}{2\sigma_o^2} (\mathbf{O}_T - M_T \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{T-1}) \mathbf{O}_{T-1} - \sum_{i=1}^4 H_{iT}^B \mathbf{O}_{iT}^B)' \\ & \quad \quad \times (\mathbf{O}_T - M_T \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{T-1}) \mathbf{O}_{T-1} - \sum_{i=1}^4 H_{iT}^B \mathbf{O}_{iT}^B) \left. \right\}\end{aligned}$$

$$\begin{aligned} &\propto \exp \left\{ -\frac{1}{2} \left[ \mathbf{O}'_T \left( \frac{1}{\sigma_z^2} K'K + \frac{1}{\sigma_o^2} I_N \right) \mathbf{O}_T \right. \right. \\ &\quad \left. \left. - 2 \left( \frac{1}{\sigma_z^2} K' \mathbf{Z}_T + \frac{1}{\sigma_o^2} (M_T \boldsymbol{\beta} + (\text{diag}(\boldsymbol{\rho}) + H_{T-1}) \mathbf{O}_{T-1} + \sum_{i=1}^4 H_{iT}^B \mathbf{O}_{iT}^B) \right)' \mathbf{O}_T \right] \right\} \end{aligned}$$

Thus

$$[\mathbf{O}_T | \cdot] \sim MVN(\Lambda_T \chi_T, \Lambda_T) \quad \text{where}$$

$$\Lambda_T^{-1} = \left( \frac{1}{\sigma_z^2} K'K + \frac{1}{\sigma_o^2} I_N \right)$$

and

$$\chi_1 = \left( \frac{1}{\sigma_z^2} K' \mathbf{Z}_T + \frac{1}{\sigma_o^2} \left( M_T \boldsymbol{\beta} + (\text{diag}(\boldsymbol{\rho}) + H_{T-1}) \mathbf{O}_{T-1} + \sum_{i=1}^4 H_{iT}^B \mathbf{O}_{iT}^B \right) \right)$$

$\mathbf{O}_{it}^B$ 's full conditional distribution follows from (3.2) and (3.4):

$$\begin{aligned} [\mathbf{O}_{it}^B | \cdot] &\propto [\mathbf{O}_{it}^B | v_{it}, \sigma_B^2] [\mathbf{O}_t | M_t, \boldsymbol{\beta}, \boldsymbol{\rho}, H_{t-1}, H_{it}^B, \mathbf{O}_{t-1}, \mathbf{O}_{it}^B, \sigma_o^2] \\ &\propto \exp \left\{ -\frac{1}{2\sigma_B^2} (\mathbf{O}_{it}^B - \mathbf{1}_{n_i} v_{it})' (\mathbf{O}_{it}^B - \mathbf{1}_{n_i} v_{it}) \right. \\ &\quad \left. - \frac{1}{2\sigma_o^2} (\mathbf{O}_t - M_t \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{t-1}) \mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B)' \right. \\ &\quad \left. \times (\mathbf{O}_t - M_t \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{t-1}) \mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \mathbf{O}_{it}^{B'} \left( \frac{1}{\sigma_B^2} I_{n_i} + \frac{1}{\sigma_o^2} H_{it}^{B'} H_{it}^B \right) \mathbf{O}_{it}^B \right. \right. \\ &\quad \left. \left. - 2 \left( \frac{1}{\sigma_B^2} \mathbf{1}'_{n_i} v_{it} + \frac{1}{\sigma_o^2} H_{it}^{B'} (\mathbf{O}_t - M_t \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{t-1}) \mathbf{O}_{t-1}) - \sum_{j \neq i} H_{it}^B \mathbf{O}_{it}^B \right)' \mathbf{O}_{it}^B \right] \right\} \end{aligned}$$

Thus

$$[\mathbf{O}_{it}^B | \cdot] \sim MVN(\Lambda_B \chi_B, \Lambda_B) \quad \text{where}$$

$$\Lambda_B^{-1} = \left( \frac{1}{\sigma_B^2} I_{n_i} + \frac{1}{\sigma_o^2} H_{it}^{B'} H_{it}^B \right)$$

and

$$\begin{aligned} \chi_B &= \left( \frac{1}{\sigma_B^2} \mathbf{1}'_{n_i} v_{it} + \frac{1}{\sigma_o^2} H_{it}^{B'} (\mathbf{O}_t - M_t \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{t-1}) \mathbf{O}_{t-1} - \sum_{j \neq i} H_{it}^B \mathbf{O}_{it}^B) \right) \\ &\quad i = 1, \dots, 4 \end{aligned}$$

$\mathbf{v}_{it}$ 's full conditional distribution follows from (3.4):

$$\begin{aligned}
[\mathbf{v}_{it} | \cdot] &\propto [\mathbf{O}_{it}^B | \mathbf{v}_{it}, \sigma_B^2][\mathbf{v}_{it} | \phi_{i1}, \phi_{i2}, \mathbf{v}_{it-1}, \sigma_{v_i}^2][\mathbf{v}_{it+1} | \phi_{i1}, \phi_{i2}, \mathbf{v}_{it}, \sigma_{v_i}^2] \\
&\propto \exp \left\{ -\frac{1}{2\sigma_B^2} (\mathbf{O}_{it}^B - \mathbf{1}_{n_i} \mathbf{v}_{it})' (\mathbf{O}_{it}^B - \mathbf{1}_{n_i} \mathbf{v}_{it}) \right. \\
&\quad -\frac{1}{2\sigma_{v_i}^2} (\mathbf{v}_{it} - \mathbf{1}_{n_i} \phi_{i1} - \phi_{i2} \mathbf{v}_{it-1})' (\mathbf{v}_{it} - \mathbf{1}_{n_i} \phi_{i1} - \phi_{i2} \mathbf{v}_{it-1}) \\
&\quad \left. -\frac{1}{2\sigma_{v_i}^2} (\mathbf{v}_{it+1} - \mathbf{1}_{n_i} \phi_{i1} - \phi_{i2} \mathbf{v}_{it})' (\mathbf{v}_{it+1} - \mathbf{1}_{n_i} \phi_{i1} - \phi_{i2} \mathbf{v}_{it}) \right. \\
&\propto \exp \left\{ -\frac{1}{2} \left[ \mathbf{v}_{it}' \left( \frac{1}{\sigma_B^2} I_{n_i} + \frac{1}{\sigma_{v_i}^2} I_{n_i} + \frac{\phi_{i2}^2}{\sigma_{v_i}^2} I_{n_i} \right) \mathbf{v}_{it} \right. \right. \\
&\quad \left. \left. -2 \left( \frac{1}{\sigma_B^2} \mathbf{O}_{it}^B + \frac{1}{\sigma_{v_i}^2} (\mathbf{1}_{n_i} \phi_{i1} + \phi_{i2} \mathbf{v}_{it-1}) + \frac{\phi_{i2}}{\sigma_{v_i}^2} (\mathbf{v}_{it+1} - \mathbf{1}_{n_i} \phi_{i1}) \right)' \mathbf{v}_{it} \right] \right\}
\end{aligned}$$

Thus

$$[\mathbf{v}_{it} | \cdot] \sim MVN(\Lambda_v \chi_v, \Lambda_v) \quad \text{where}$$

$$\Lambda_v^{-1} = \left( \frac{1}{\sigma_B^2} I_{n_i} + \frac{1}{\sigma_{v_i}^2} I_{n_i} + \frac{\phi_{i2}^2}{\sigma_{v_i}^2} I_{n_i} \right)$$

and

$$\chi_v = \left( \frac{1}{\sigma_B^2} \mathbf{O}_{it}^B + \frac{1}{\sigma_{v_i}^2} (\mathbf{1}_{n_i} \phi_{i1} + \phi_{i2} \mathbf{v}_{it-1}) + \frac{\phi_{i2}}{\sigma_{v_i}^2} (\mathbf{v}_{it+1} - \mathbf{1}_{n_i} \phi_{i1}) \right)$$

$$i = 1, \dots, 4; \quad t = 2, \dots, T-1$$

For  $t = 1$  we have

$$\begin{aligned}
[\mathbf{v}_{i1} | \cdot] &\propto [\mathbf{O}_{i1}^B | \mathbf{v}_{i1}, \sigma_B^2][\mathbf{v}_{i1} | \phi_{i0}, \sigma_{v_{i1}}^2][\mathbf{v}_{i2} | \phi_{i1}, \phi_{i2}, \mathbf{v}_{i1}, \sigma_{v_i}^2] \\
&\propto \exp \left\{ -\frac{1}{2\sigma_B^2} (\mathbf{O}_{i1}^B - \mathbf{1}_{n_i} \mathbf{v}_{i1})' (\mathbf{O}_{i1}^B - \mathbf{1}_{n_i} \mathbf{v}_{i1}) \right. \\
&\quad -\frac{1}{2\sigma_{v_{i1}}^2} (\mathbf{v}_{i1} - \mathbf{1}_{n_i} \phi_{i0})' (\mathbf{v}_{i1} - \mathbf{1}_{n_i} \phi_{i0}) \\
&\quad \left. -\frac{1}{2\sigma_{v_i}^2} (\mathbf{v}_{i2} - \mathbf{1}_{n_i} \phi_{i1} - \phi_{i2} \mathbf{v}_{i1})' (\mathbf{v}_{i2} - \mathbf{1}_{n_i} \phi_{i1} - \phi_{i2} \mathbf{v}_{i1}) \right. \\
&\propto \exp \left\{ -\frac{1}{2} \left[ \mathbf{v}_{i1}' \left( \frac{1}{\sigma_B^2} I_{n_i} + \frac{1}{\sigma_{v_{i1}}^2} I_{n_i} + \frac{\phi_{i2}^2}{\sigma_{v_i}^2} I_{n_i} \right) \mathbf{v}_{i1} \right. \right. \\
&\quad \left. \left. -2 \left( \frac{1}{\sigma_B^2} \mathbf{O}_{i1}^B + \frac{1}{\sigma_{v_{i1}}^2} \mathbf{1}_{n_i} \phi_{i0} + \frac{\phi_{i2}}{\sigma_{v_i}^2} (\mathbf{v}_{i2} - \mathbf{1}_{n_i} \phi_{i1}) \right)' \mathbf{v}_{i1} \right] \right\}
\end{aligned}$$

Thus

$[\mathbf{v}_{i1} | \cdot] \sim MVN(\Lambda_{v1}\chi_{v1}, \Lambda_{v1})$  where

$$\Lambda_{v1}^{-1} = \left( \frac{1}{\sigma_B^2} I_{n_i} + \frac{1}{\sigma_{v_{0i}}^2} I_{n_i} + \frac{\phi_{i2}^2}{\sigma_{v_i}^2} I_{n_i} \right)$$

and

$$\chi_{v1} = \left( \frac{\mathbf{1}}{\sigma_B^2} \mathbf{O}_{i1}^B + \frac{1}{\sigma_{v_{0i}}^2} \mathbf{1}_{n_i} \phi_{i0} + \frac{\phi_{i2}}{\sigma_{v_i}^2} (\mathbf{v}_{i2} - \mathbf{1}_{n_i} \phi_{i1}) \right)$$

$$i = 1, \dots, 4$$

Similar for  $t = T$

$$\begin{aligned} [\mathbf{v}_{iT} | \cdot] &\propto [\mathbf{O}_{iT}^B | \mathbf{v}_{iT}, \sigma_B^2] [\mathbf{v}_{iT} | \phi_{i1}, \phi_{i2}, \mathbf{v}_{iT}, \sigma_{v_i}^2] \\ &\propto \exp \left\{ -\frac{1}{2\sigma_B^2} (\mathbf{O}_{iT}^B - \mathbf{1}_{n_i} \mathbf{v}_{iT})' (\mathbf{O}_{iT}^B - \mathbf{1}_{n_i} \mathbf{v}_{iT}) \right. \\ &\quad \left. - \frac{1}{2\sigma_{v_i}^2} (\mathbf{v}_{iT} - \mathbf{1}_{n_i} \phi_{i1} - \phi_{i2} \mathbf{v}_{iT-1})' (\mathbf{v}_{iT} - \mathbf{1}_{n_i} \phi_{i1} - \phi_{i2} \mathbf{v}_{iT-1}) \right. \\ &\quad \left. \exp \left\{ -\frac{1}{2} \left[ \mathbf{v}_{iT}' \left( \frac{1}{\sigma_B^2} I_{n_i} + \frac{1}{\sigma_{v_i}^2} I_{n_i} \right) \mathbf{v}_{iT} - 2 \left( \frac{1}{\sigma_B^2} \mathbf{O}_{iT}^B + \frac{1}{\sigma_{v_i}^2} (\mathbf{1}_{n_i} \phi_{i1} + \phi_{i2} \mathbf{v}_{iT-1}) \right)' \mathbf{v}_{iT} \right] \right\} \right\} \end{aligned}$$

Thus

$[\mathbf{v}_{iT} | \cdot] \sim MVN(\Lambda_{vT}\chi_{vT}, \Lambda_{vT})$  where

$$\Lambda_{vT}^{-1} = \left( \frac{1}{\sigma_B^2} I_{n_i} + \frac{1}{\sigma_{v_i}^2} I_{n_i} \right)$$

and

$$\chi_{vT} = \left( \frac{\mathbf{1}}{\sigma_B^2} \mathbf{O}_{iT}^B + \frac{1}{\sigma_{v_i}^2} (\mathbf{1}_{n_i} \phi_{i1} + \phi_{i2} \mathbf{v}_{iT-1}) \right)$$

$$i = 1, \dots, 4$$

$\phi_{i\cdot}$ 's full conditional distribution follows from (3.4) and (3.6):

$$[\phi_{i1} | \cdot] \propto [\phi_{i1} | \tilde{\phi}_{i1}, \tilde{\sigma}_\phi^2] \prod_{t=2}^T [\mathbf{v}_{it} | \phi_{i1}, \phi_{i2}, \mathbf{v}_{it-1}, \sigma_{v_i}^2]$$

$$\propto \exp\left\{-\frac{1}{2\tilde{\sigma}_\phi^2}(\phi_{i1} - \tilde{\phi}_{i1})^2 - \frac{1}{2\sigma_{v_i}^2}\sum_{t=2}^T(\mathbf{v}_{it} - \mathbf{1}_{n_i}\phi_{i1} - \phi_{i2}\mathbf{v}_{it-1})'(\mathbf{v}_{it} - \mathbf{1}_{n_i}\phi_{i1} - \phi_{i2}\mathbf{v}_{it-1})\right\}$$

Thus

$$[\phi_{i1} | \cdot] \sim N(\Lambda_{\phi_1}\chi_{\phi_1}, \Lambda_{\phi_1})$$

$$\Lambda_{\phi_1}^{-1} = \left(\frac{1}{\tilde{\sigma}_\phi^2} + \frac{1}{\sigma_{v_i}^2}\sum_{t=2}^T\mathbf{1}'_{n_i}\mathbf{1}_{n_i}\right)$$

and

$$\chi_{\phi_1} = \left(\frac{1}{\tilde{\sigma}_\phi^2}\tilde{\phi}_{i1} + \frac{1}{\sigma_{v_i}^2}\sum_{t=2}^T\mathbf{1}'_{n_i}(\mathbf{v}_{it} - \phi_{i2}\mathbf{v}_{it-1})\right)$$

$$i = 1, \dots, 4$$

$$\begin{aligned} [\phi_{i2} | \cdot] &\propto [\phi_{i2} | \tilde{\phi}_{i2}, \tilde{\sigma}_\phi^2] \prod_{t=2}^T [\mathbf{v}_{it} | \phi_{i1}, \phi_{i2}, \mathbf{v}_{it-1}, \sigma_{v_i}^2] \\ &\propto \exp\left\{-\frac{1}{2\tilde{\sigma}_\phi^2}(\phi_{i2} - \tilde{\phi}_{i2})^2 - \frac{1}{2\sigma_{v_i}^2}\sum_{t=2}^T(\mathbf{v}_{it} - \mathbf{1}_{n_i}\phi_{i1} - \phi_{i2}\mathbf{v}_{it-1})'(\mathbf{v}_{it} - \mathbf{1}_{n_i}\phi_{i1} - \phi_{i2}\mathbf{v}_{it-1})\right\} \end{aligned}$$

Thus

$$[\phi_{i2} | \cdot] \sim N(\Lambda_{\phi_2}\chi_{\phi_2}, \Lambda_{\phi_2})$$

$$\Lambda_{\phi_2}^{-1} = \left(\frac{1}{\tilde{\sigma}_\phi^2} + \frac{1}{\sigma_{v_i}^2}\sum_{t=2}^T\mathbf{v}'_{it-1}\mathbf{v}_{it-1}\right)$$

and

$$\chi_{\phi_2} = \left(\frac{1}{\tilde{\sigma}_\phi^2}\tilde{\phi}_{i2} + \frac{1}{\sigma_{v_i}^2}\sum_{t=2}^T\mathbf{v}'_{it-1}(\mathbf{v}_{it} - \mathbf{1}_{n_i}\phi_{i1})\right)$$

$$i = 1, \dots, 4$$

$$\begin{aligned} [\phi_{i0} | \cdot] &\propto [\phi_{i0} | \tilde{\phi}_{i0}, \tilde{\sigma}_\phi^2][\mathbf{v}_{i1} | \phi_{i0}, \sigma_{v_{0i}}^2] \\ &\propto \exp\left\{-\frac{1}{2\tilde{\sigma}_\phi^2}(\phi_{i0} - \tilde{\phi}_{i0})^2 - \frac{1}{2\sigma_{v_{0i}}^2}(\mathbf{v}_{i1} - \mathbf{1}_{n_i}\phi_{i0})'(\mathbf{v}_{i1} - \mathbf{1}_{n_i}\phi_{i0})\right\} \end{aligned}$$

Thus

$$[\phi_{i0} | \cdot] \sim N\left(\left(\frac{1}{\tilde{\sigma}_\phi^2} + \frac{1}{\sigma_{v_{0i}}^2}\right)^{-1} \left(\frac{1}{\tilde{\sigma}_\phi^2}\tilde{\phi}_{i0} + \frac{1}{\sigma_{v_{0i}}^2}\mathbf{1}'_{n_i}\mathbf{v}_{i1}\right), \left(\frac{1}{\tilde{\sigma}_\phi^2} + \frac{1}{\sigma_{v_{0i}}^2}\right)^{-1}\right)$$

$$i = 1, \dots, 4$$

$\boldsymbol{\beta}$ 's full conditional distribution follows from (2.20) and (3.2):

$$\begin{aligned} [\boldsymbol{\beta} | \cdot] &\propto [\boldsymbol{\beta} | \tilde{\boldsymbol{\beta}}, \tilde{\sigma}_\beta^2] [\mathbf{O}_1 | M_1, \boldsymbol{\beta}, H_{i1}^B, \mathbf{O}_{i1}^B, \sigma_o^2] \prod_{t=2}^T [\mathbf{O}_t | M_t, \boldsymbol{\beta}, \boldsymbol{\rho}, H_{t-1}, H_{it}^B, \mathbf{O}_{t-1}, \mathbf{O}_{it}^B, \sigma_o^2] \\ &\propto \exp \left\{ -\frac{1}{2\tilde{\sigma}_\beta^2} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})' (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) \right. \\ &\quad - \frac{1}{2\sigma_o^2} (\mathbf{O}_1 - M_1 \boldsymbol{\beta} - \sum_{i=1}^4 H_{i1}^B \mathbf{O}_{i1}^B)' (\mathbf{O}_1 - M_1 \boldsymbol{\beta} - \sum_{i=1}^4 H_{i1}^B \mathbf{O}_{i1}^B) \\ &\quad - \frac{1}{2\sigma_o^2} \sum_{t=2}^T (\mathbf{O}_t - M_t \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{t-1}) \mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B)' \\ &\quad \quad \times (\mathbf{O}_t - M_t \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{t-1}) \mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B) \left. \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\beta}' \left( \frac{1}{\tilde{\sigma}_\beta^2} + \frac{1}{\sigma_o^2} \sum_{t=1}^T M_t' M_t \right) \boldsymbol{\beta} - 2 \left( \frac{1}{\tilde{\sigma}_\beta^2} \tilde{\boldsymbol{\beta}} + \frac{1}{\sigma_o^2} M_1' (\mathbf{O}_1 - \sum_{i=1}^4 H_{i1}^B \mathbf{O}_{i1}^B) \right. \right. \right. \\ &\quad \left. \left. \left. + \frac{1}{\sigma_o^2} \sum_{t=2}^T M_t' (\mathbf{O}_t - (\text{diag}(\boldsymbol{\rho}) + H_{t-1}) \mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B) \right)' \boldsymbol{\beta} \right] \right\} \end{aligned}$$

Thus

$$[\boldsymbol{\beta} | \cdot] \sim N(\Lambda_\beta \chi_\beta, \Lambda_\beta)$$

$$\Lambda_\beta^{-1} = \left( \frac{1}{\tilde{\sigma}_\beta^2} + \frac{1}{\sigma_o^2} \sum_{t=1}^T M_t' M_t \right)$$

and

$$\begin{aligned} \chi_\beta = \left( \frac{1}{\tilde{\sigma}_\beta^2} \tilde{\boldsymbol{\beta}} + \frac{1}{\sigma_o^2} M_1' (\mathbf{O}_1 - \sum_{i=1}^4 H_{i1}^B \mathbf{O}_{i1}^B) + \frac{1}{\sigma_o^2} \sum_{t=2}^T M_t' (\mathbf{O}_t - (\text{diag}(\boldsymbol{\rho}) + \right. \\ \left. H_{t-1}) \mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B) \right) \end{aligned}$$

$\boldsymbol{\rho}$ 's full conditional distribution follows from (2.20) and (3.2):

$$\begin{aligned} [\boldsymbol{\rho} | \cdot] &\propto [\boldsymbol{\rho} | \tilde{\boldsymbol{\rho}}, \tilde{\sigma}_\rho^2] \prod_{t=2}^T [\mathbf{O}_t | M_t, \boldsymbol{\beta}, \boldsymbol{\rho}, H_{t-1}, H_{it}^B, \mathbf{O}_{t-1}, \mathbf{O}_{it}^B, \sigma_o^2] \\ &\propto \exp \left\{ -\frac{1}{2\tilde{\sigma}_\rho^2} (\boldsymbol{\rho} - \tilde{\boldsymbol{\rho}})' (\boldsymbol{\rho} - \tilde{\boldsymbol{\rho}}) \right. \\ &\quad - \frac{1}{2\sigma_o^2} \sum_{t=2}^T (\mathbf{O}_t - M_t \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{t-1}) \mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B)' \\ &\quad \quad \times (\mathbf{O}_t - M_t \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{t-1}) \mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B) \left. \right\} \end{aligned}$$

$$\propto \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\rho}' \left( \frac{1}{\tilde{\sigma}_\rho^2} + \frac{1}{\sigma_0^2} \sum_{t=2}^T \mathbf{O}'_{t-1} \mathbf{O}_{t-1} \right) \boldsymbol{\rho} - 2 \left( \frac{1}{\tilde{\sigma}_\rho^2} \tilde{\boldsymbol{\rho}} + \frac{1}{\sigma_0^2} \sum_{t=2}^T \mathbf{O}'_{t-1} (\mathbf{O}_t - M_t \boldsymbol{\beta} - H_{t-1} \mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B) \right)' \boldsymbol{\rho} \right] \right\}$$

Thus

$$[\boldsymbol{\rho} | \cdot] \sim N(\Lambda_\rho \chi_\rho, \Lambda_\rho)$$

$$\Lambda_\rho^{-1} = \left( \frac{1}{\tilde{\sigma}_\rho^2} + \frac{1}{\sigma_0^2} \sum_{t=2}^T \mathbf{O}'_{t-1} \mathbf{O}_{t-1} \right)$$

and

$$\chi_\rho = \left( \frac{1}{\tilde{\sigma}_\rho^2} \tilde{\boldsymbol{\rho}} + \frac{1}{\sigma_0^2} \sum_{t=2}^T \mathbf{O}'_{t-1} (\mathbf{O}_t - M_t \boldsymbol{\beta} - H_{t-1} \mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B) \right)$$

$\boldsymbol{\theta}$ 's full conditional distribution follows from (2.20) and (3.2):

$$\begin{aligned} [\boldsymbol{\theta} | \cdot] &\propto [\boldsymbol{\theta} | \tilde{\boldsymbol{\theta}}, \tilde{\sigma}_\theta^2] [\mathbf{O}_1 | M_1, \boldsymbol{\beta}, G_{11}^B, \boldsymbol{\theta}, \sigma_0^2] \prod_{t=2}^T [\mathbf{O}_t | M_t, \boldsymbol{\beta}, \boldsymbol{\rho}, G_{t-1}, G_{it}^B, \boldsymbol{\theta}, \mathbf{O}_{t-1}, \sigma_0^2] \\ &\propto \exp \left\{ -\frac{1}{2\tilde{\sigma}_\theta^2} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})' (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) \right. \\ &\quad - \frac{1}{2\sigma_0^2} (\mathbf{O}_1 - M_1 \boldsymbol{\beta} - \sum_{i=1}^4 G_{i1}^B \boldsymbol{\theta})' (\mathbf{O}_1 - M_1 \boldsymbol{\beta} - \sum_{i=1}^4 G_{i1}^B \boldsymbol{\theta}) \\ &\quad - \frac{1}{2\sigma_0^2} \sum_{t=2}^T (\mathbf{O}_t - M_t \boldsymbol{\beta} - \text{diag}(\boldsymbol{\rho}) \mathbf{O}_{t-1} - G_{t-1} \boldsymbol{\theta} - \sum_{i=1}^4 G_{it}^B \boldsymbol{\theta})' \\ &\quad \quad \times (\mathbf{O}_t - M_t \boldsymbol{\beta} - \text{diag}(\boldsymbol{\rho}) \mathbf{O}_{t-1} - G_{t-1} \boldsymbol{\theta} - \sum_{i=1}^4 G_{it}^B \boldsymbol{\theta}) \left. \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\theta}' \left( \frac{1}{\tilde{\sigma}_\theta^2} + \frac{1}{\sigma_0^2} \sum_{i,j}^4 G_{i1}^{B'} G_{j1}^B + \frac{1}{\sigma_0^2} \sum_{t=2}^T \left( (G_{t-1} + \sum_{i,j}^4 G_{jt}^B)' (G_{t-1} + \sum_{i,j}^4 G_{jt}^B) \right) \right) \boldsymbol{\theta} \right. \right. \\ &\quad - 2 \left( \frac{1}{\tilde{\sigma}_\theta^2} \tilde{\boldsymbol{\theta}} + \frac{1}{\sigma_0^2} (\sum_{i=1}^4 G_{i1}^B)' (\mathbf{O}_1 - M_1 \boldsymbol{\beta}) \right. \\ &\quad \quad \left. \left. + \sum_{t=2}^T (G_{t-1} + \sum_{i=1}^4 G_{it}^B)' (\mathbf{O}_t - M_t \boldsymbol{\beta} - \text{diag}(\boldsymbol{\rho}) \mathbf{O}_{t-1}) \right)' \boldsymbol{\theta} \right] \left. \right\} \end{aligned}$$

in the above presentation of  $[\mathbf{O}_t | \cdot]$  we use design matrixes  $G_t$  and  $G_t^B$  instead of the parametric matrixes  $H_t$  and  $H_t^B$ , i.e.  $G(\mathbf{O}_t; \mathbf{U}_t, \mathbf{V}_t) \boldsymbol{\theta} = H(\boldsymbol{\theta}; \mathbf{U}_t, \mathbf{V}_t) \mathbf{O}_t$ .

Thus

$$[\boldsymbol{\theta} | \cdot] \sim N(\Lambda_\theta \chi, \Lambda_\theta)$$

$$\Lambda_\theta^{-1} = \left( \frac{1}{\tilde{\sigma}_\theta^2} + \frac{1}{\sigma_0^2} \sum_{i,j}^4 G_{i1}^{B'} G_{j1}^B + \frac{1}{\sigma_0^2} \sum_{t=2}^T \left( (G_{t-1} + \sum_{i,j}^4 G_{jt}^B)' (G_{t-1} + \sum_{i,j}^4 G_{jt}^B) \right) \right)$$

and

$$\chi = \left( \frac{1}{\sigma_\theta^2} \tilde{\boldsymbol{\theta}} + \frac{1}{\sigma_o^2} (\sum_{i=1}^4 G_{i1}^B)' (\mathbf{O}_1 - M_1 \boldsymbol{\beta}) + \sum_{t=2}^T (G_{t-1} + \sum_{i=1}^4 G_{it}^B)' (\mathbf{O}_t - M_t \boldsymbol{\beta} - \text{diag}(\boldsymbol{\rho}) \mathbf{O}_{t-1}) \right)$$

$\sigma_z^2$ 's full conditional distribution follows from (2.1) and (2.20):

$$\begin{aligned} [\sigma_z^2 | \cdot] &\propto [\sigma_z^2 | \tilde{a}_z, \tilde{b}_z] \prod_{t=1}^T [\mathbf{Z}_t | K, \mathbf{O}_t, \sigma_z^2] \\ &\propto \frac{1}{\Gamma(\tilde{a}_z)} \tilde{b}_z^{-\tilde{a}_z} (\sigma_z^2)^{-\tilde{a}_z-1} \exp\left\{-\frac{\tilde{b}_z}{\sigma_z^2}\right\} (\sigma_z^2)^{-\frac{nT}{2}} \exp\left\{-\frac{1}{2\sigma_z^2} \sum_{t=1}^T (\mathbf{Z}_t - K \mathbf{O}_t)' (\mathbf{Z}_t - K \mathbf{O}_t)\right\} \\ &\propto (\sigma_z^2)^{-\frac{nT}{2} - \tilde{a}_z - 1} \exp\left\{-\frac{1}{\sigma_z^2} \left[ \tilde{b}_z + \frac{1}{2} \sum_{t=1}^T (\mathbf{Z}_t - K \mathbf{O}_t)' (\mathbf{Z}_t - K \mathbf{O}_t) \right]\right\} \end{aligned}$$

Thus

$$[\sigma_z^2 | \cdot] \propto \mathbf{IG} \left( \frac{nT}{2} + \tilde{a}_z, \tilde{b}_z + \frac{1}{2} \sum_{t=1}^T (\mathbf{Z}_t - K \mathbf{O}_t)' (\mathbf{Z}_t - K \mathbf{O}_t) \right)$$

$\sigma_o^2$ 's full conditional distribution follows from (2.1) and (2.20):

$$\begin{aligned} [\sigma_o^2 | \cdot] &\propto [\sigma_o^2 | \tilde{a}_o, \tilde{b}_o] [\mathbf{O}_1 | M_1, \boldsymbol{\beta}, G_{11}^B, \boldsymbol{\theta}, \sigma_o^2] \prod_{t=2}^T [\mathbf{O}_t | M_t, \boldsymbol{\beta}, \boldsymbol{\rho}, G_{t-1}, G_{it}^B, \boldsymbol{\theta}, \mathbf{O}_{t-1}, \sigma_o^2] \\ &\propto \frac{1}{\Gamma(\tilde{a}_o)} \tilde{b}_o^{-\tilde{a}_o} (\sigma_o^2)^{-\tilde{a}_o-1} \exp\left\{-\frac{\tilde{b}_o}{\sigma_o^2}\right\} \\ &\quad \times (\sigma_o^2)^{-\frac{NT}{2}} \exp\left\{-\frac{1}{2\sigma_o^2} (\mathbf{O}_1 - M_1 \boldsymbol{\beta} - \sum_{i=1}^4 H_{i1}^B \mathbf{O}_{i1}^B)' (\mathbf{O}_1 - M_1 \boldsymbol{\beta} - \sum_{i=1}^4 H_{i1}^B \mathbf{O}_{i1}^B) \right. \\ &\quad \left. - \frac{1}{2\sigma_o^2} \sum_{t=2}^T (\mathbf{O}_t - M_t \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{t-1}) \mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B)' \right. \\ &\quad \left. \times (\mathbf{O}_t - M_t \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{t-1}) \mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B) \right\} \end{aligned}$$

Thus

$$\begin{aligned} [\sigma_o^2 | \cdot] &\propto \mathbf{IG} \left( \frac{NT}{2} + \tilde{a}_o, \tilde{b}_o + \frac{1}{2} \left[ (\mathbf{O}_1 - M_1 \boldsymbol{\beta} - \sum_{i=1}^4 H_{i1}^B \mathbf{O}_{i1}^B)' (\mathbf{O}_1 - M_1 \boldsymbol{\beta} - \sum_{i=1}^4 H_{i1}^B \mathbf{O}_{i1}^B) \right. \right. \\ &\quad \left. \left. + \sum_{t=2}^T (\mathbf{O}_t - M_t \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{t-1}) \mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B)' \right. \right. \\ &\quad \left. \left. \times (\mathbf{O}_t - M_t \boldsymbol{\beta} - (\text{diag}(\boldsymbol{\rho}) + H_{t-1}) \mathbf{O}_{t-1} - \sum_{i=1}^4 H_{it}^B \mathbf{O}_{it}^B) \right] \right) \end{aligned}$$



$\sigma_B^2$ 's full conditional distribution follows from (3.4) and (3.5):

$$\begin{aligned}
[\sigma_B^2 | \cdot] &\propto [\sigma_B^2 | \tilde{a}_B, \tilde{b}_B] \prod_{t=1}^T \prod_{i=1}^4 [\mathbf{O}_{it}^B | \mathbf{v}_{it}, \sigma_B^2] \\
&\propto \frac{1}{\Gamma(\tilde{a}_B)} \tilde{b}_B^{-\tilde{a}_B} (\sigma_B^2)^{-\tilde{a}_B-1} \exp\left\{-\frac{\tilde{b}_B}{\sigma_B^2}\right\} \\
&\quad \times (\sigma_B^2)^{-\frac{T \sum_{i=1}^4 n_i}{2}} \exp\left\{-\frac{1}{2\sigma_B^2} \sum_{t=1}^T \sum_{i=1}^4 (\mathbf{O}_{it}^B - \mathbf{1}_{n_i} \mathbf{v}_{it})' (\mathbf{O}_{it}^B - \mathbf{1}_{n_i} \mathbf{v}_{it})\right\} \\
&\propto (\sigma_B^2)^{-\frac{T \sum_{i=1}^4 n_i}{2} - \tilde{a}_B - 1} \exp\left\{-\frac{1}{\sigma_B^2} \left[ \tilde{b}_B + \frac{1}{2} \sum_{t=1}^T \sum_{i=1}^4 (\mathbf{O}_{it}^B - \mathbf{1}_{n_i} \mathbf{v}_{it})' (\mathbf{O}_{it}^B - \mathbf{1}_{n_i} \mathbf{v}_{it}) \right]\right\}
\end{aligned}$$

Thus

$$[\sigma_B^2 | \cdot] \propto \mathbf{IG} \left( \frac{T \sum_{i=1}^4 n_i}{2} + \tilde{a}_B, \tilde{b}_B + \frac{1}{2} \sum_{t=1}^T \sum_{i=1}^4 (\mathbf{O}_{it}^B - \mathbf{1}_{n_i} \mathbf{v}_{it})' (\mathbf{O}_{it}^B - \mathbf{1}_{n_i} \mathbf{v}_{it}) \right)$$

$\sigma_{v_i}^2$ 's full conditional distribution follows from (3.4) and (3.5):

$$\begin{aligned}
[\sigma_{v_i}^2 | \cdot] &\propto [\sigma_{v_i}^2 | \tilde{a}_v, \tilde{b}_v] \prod_{t=2}^T [\mathbf{v}_{it} | \phi_{i1}, \phi_{i2}, \mathbf{v}_{it-1}, \sigma_{v_i}^2] \\
&\propto \frac{1}{\Gamma(\tilde{a}_v)} \tilde{b}_v^{-\tilde{a}_v} (\sigma_{v_i}^2)^{-\tilde{a}_v-1} \exp\left\{-\frac{\tilde{b}_v}{\sigma_{v_i}^2}\right\} \\
&\quad \times (\sigma_{v_i}^2)^{-\frac{T n_i}{2}} \exp\left\{-\frac{1}{2\sigma_{v_i}^2} \sum_{t=2}^T (\mathbf{v}_{it} - \mathbf{1}_{n_i} \phi_{i1} - \phi_{i2} \mathbf{v}_{it-1})' (\mathbf{v}_{it} - \mathbf{1}_{n_i} \phi_{i1} - \phi_{i2} \mathbf{v}_{it-1})\right\} \\
&\propto (\sigma_{v_i}^2)^{-\frac{T n_i}{2} - \tilde{a}_v - 1} \exp\left\{-\frac{1}{\sigma_{v_i}^2} \left[ \tilde{b}_v + \frac{1}{2} \sum_{t=2}^T (\mathbf{v}_{it} - \mathbf{1}_{n_i} \phi_{i1} - \phi_{i2} \mathbf{v}_{it-1})' \right. \right. \\
&\quad \left. \left. \times (\mathbf{v}_{it} - \mathbf{1}_{n_i} \phi_{i1} - \phi_{i2} \mathbf{v}_{it-1}) \right]\right\}
\end{aligned}$$

Thus

$$[\sigma_{v_i}^2 | \cdot] \propto \mathbf{IG} \left( \frac{T n_i}{2} + \tilde{a}_v, \tilde{b}_v + \sum_{t=2}^T (\mathbf{v}_{it} - \mathbf{1}_{n_i} \phi_{i1} - \phi_{i2} \mathbf{v}_{it-1})' (\mathbf{v}_{it} - \mathbf{1}_{n_i} \phi_{i1} - \phi_{i2} \mathbf{v}_{it-1}) \right)$$

$i = 1, \dots, 4$

$\sigma_{v_{0i}}^2$ 's full conditional distribution follows from (3.4) and (3.5):

$$\begin{aligned}
[\sigma_{v_{0i}}^2 | \cdot] &\propto [\sigma_{v_{0i}}^2 | \tilde{a}_v, \tilde{b}_v][v_{i1} | \phi_{i0}, \sigma_{v_{0i}}^2] \\
&\propto \frac{1}{\Gamma(\tilde{a}_v)} \tilde{b}_v^{-\tilde{a}_v} (\sigma_{v_{0i}}^2)^{-\tilde{a}_v-1} \exp\left\{-\frac{\tilde{b}_v}{\sigma_{v_{0i}}^2}\right\} \\
&\quad \times (\sigma_{v_{0i}}^2)^{-\frac{n_i}{2}} \exp\left\{-\frac{1}{2\sigma_{v_{0i}}^2} (\mathbf{v}_{i1} - \mathbf{1}_{n_i} \phi_{i0})' (\mathbf{v}_{i1} - \mathbf{1}_{n_i} \phi_{i0})\right\} \\
&\propto (\sigma_{v_{0i}}^2)^{-\frac{n_i}{2}-\tilde{a}_v-1} \exp\left\{-\frac{1}{\sigma_{v_{0i}}^2} \left[\tilde{b}_v + \frac{1}{2} (\mathbf{v}_{i1} - \mathbf{1}_{n_i} \phi_{i0})' (\mathbf{v}_{i1} - \mathbf{1}_{n_i} \phi_{i0})\right]\right\}
\end{aligned}$$

Thus

$$[\sigma_{v_{0i}}^2 | \cdot] \propto \mathbf{IG}\left(\frac{n_i}{2} + \tilde{a}_v, \tilde{b}_v + (\mathbf{v}_{i1} - \mathbf{1}_{n_i} \phi_{i0})' (\mathbf{v}_{i1} - \mathbf{1}_{n_i} \phi_{i0})\right); \quad i = 1, \dots, 4$$

## APPENDIX B

### Extension to three of data

This section provides model results and summaries of the observed ozone and meteorological fields for the years 2004, 2005 and 2006. In Chapter 4 we presented model results for the year 2004 only. However, we broaden the analysis and apply the model for two more years: 2005 and 2006. The extended period includes one year of high (2006), one of low (2005) and one of intermediate (2004) ozone concentrations, as summarized in Table B.2. Tables B.1 provide summary statistics for meteorological conditions during ozone season in all three years.

Tables B.3 to B.6 provide summaries of model performance and parameters in all three years, following from Table 4.1 to 4.4.

**Table B.1** Meteorological data summaries for three years studied

Data summaries	2004		2005		2006	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Temperature	19.87	0.8	19.36	0.8	19.56	0.9
Relative Humidity	58.5	3.1	57.73	3.7	53.48	3.8
Solar Radiation	407.68	25.3	387.34	24.1	416.52	24.9
Wind Speed East-West	5.37	3.3	5.28	3.4	5.88	3.1
Wind Speed North-South	2.21	2.1	2.18	1.8	2.14	2.0

**Table B.2** Ozone summaries for three years studied

Percentiles	2004	2005	2006
50%	29.7	27.4	29.7
75%	36.6	33.8	36.5
90%	42.8	39.8	43.5
95%	50.7	44.1	48.1
97.5%	57.4	49.5	54.8

**Table B.3** Model  $Model_1$  parameter and performance statistics summaries for three years studied.

Parameter	2004		2005		2006	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
$\beta_1$	0.429	0.03	0.391	0.03	0.145	0.03
$\beta_2$	-0.135	0.01	-0.121	0.001	-0.097	0.01
$\beta_3$	0.001	0.001	0.001	0.001	0.007	0.001
$\theta_1$	0.003	0.001	0.006	0.001	0.007	0.001
$\theta_2$	0.001	0.001	0.003	0.001	0.001	0.01
$\rho$	0.54	0.02	0.55	0.01	0.619	0.02
$\sigma_z^2$	33.33	1.33	25.92	1.024	45.15	1.47
$\sigma_o^2$	24.1	1.46	17.93	1.063	13.16	1.04

Model	2004	2005	2006
RMSE	4.26	3.802	5.458
RMSPE	6.58	5.79	6.633
Tr.	1%	1%	1%

**Table B.4** Model  $Model_2$  parameter and performance statistics summaries for three years studied.

Parameter	2004		2005		2006	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
$\beta_1$	0.462	0.027	0.388	0.035	0.179	0.033
$\beta_2$	-0.141	0.005	-0.124	0.006	-0.107	0.006
$\beta_3$	-0.002	0.001	0.001	0.001	0.007	0.001
$\theta_1$	0.017	0.002	-0.001	0.004	0.008	0.002
$\theta_2$	0.011	0.002	0.001	0.008	0.014	0.003
$\sigma_z^2$	35.206	3.795	27.604	3.445	47.682	2.826
$\sigma_o^2$	18.43	1.118	16.567	1.211	10.255	0.946

Model performance	2004	2005	2006
RMSE	4.48	3.91	5.73
RMSPE	6.32	5.75	6.68
Tr.	8%	1.5%	2.1%

**Table B.5** Model  $Model_3$  parameter and performance statistics summaries for three years studied.

Parameter	2004		2005		2006	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
$\beta_1$	0.463	0.03	0.571	0.03	0.191	0.03
$\beta_2$	-0.143	0.001	-0.167	0.01	-0.107	0.01
$\beta_3$	0.001	0.001	-0.002	0.001	0.006	0.001
$\theta_1$	0.009	0.001	0.011	0.001	0.005	0.001
$\theta_2$	0.004	0.001	0.01	0.001	0.009	0.001
$\sigma_z^2$	33.70	6.51	25.10	8.907	46.55	6.378
$\sigma_o^2$	22.56	1.09	28.46	1.597	10.84	0.897
$\phi_r$	(0.02,0.9,0.9)	(.03, .06, .18)	(.02, .93,1.0)	(.01, .03, .2)	(.10, .741, .9)	(.13, .31, .18)
$\phi_c$	(-0.01, 0.9, 1)	(.001,.02,.18)	(.01, .98, 1.0)	(.01, .02, .2)	(0.01,1.0, 0.9)	(.01, .02, .18)
$(\sigma_r^2, \sigma_{r0}^2)$	(0.001, 0.07)	(0.03, 0.08)	(0.007, 0.06)	(0.03, 0.09)	(0.001, 0.07)	(0.02, 0.08)
$(\sigma_c^2, \sigma_{c0}^2)$	(0.003, 0.067)	(0.1, 0.09)	(0.004, 0.06)	(0.1, 0.08)	(0.002, 0.068)	(0.07, 0.09)
$\sigma_b^2$	10.91	3.53	2.97	2.3	2.10	2.38

Model performance	2004	2005	2006
RMSE	4.29	3.36	5.68
RMSPE	6.19	6.27	6.69
Tr.	3.3%	5.3%	2%

**Table B.6** Model  $Model_4$  parameter and performance statistics summaries for three years studied.

Parameter	2004		2005		2006	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
$\beta_1$	0.585	0.034	0.379	0.03	0.159	0.02
$\beta_2$	-0.166	0.006	-0.12	0.001	-0.1	0.001
$\beta_3$	-0.003	0.001	0.001	0.001	0.007	0.001
$\theta_1$	-0.004	0.003	0.001	0.001	0.004	0.001
$\theta_2$	0.001	0.005	0.001	0.001	0.002	0.001
$\rho$	0.479	0.016	0.546	0.019	0.61	0.014
$\sigma_z^2$	24.86	6.116	25.63	1.61	44.54	2.073
$\sigma_o^2$	30.61	1.503	18.89	1.02	14.25	0.925
$\phi_r$	(2.53,-0.34, 1)	(1.16, .44,.22)	(.012,.96,.9)	(.01,.02,.18)	(0.01,0.96, 1)	(.01, .02, .18)
$\phi_c$	(0.35, 0.12, 1)	(.34, .53, .25)	(.001,.98,.99)	(0, 0.02,0.19)	(0.001,0.99,1)	(.001,.01,.18)
$(\sigma_r^2, \sigma_{r0}^2)$	(0.64, 0.07)	(1.34, 0.08)	(0.001, 0.07)	(0.02, 0.08)	(0.001, 0.07)	(0.02, 0.07)
$(\sigma_c^2, \sigma_{c0}^2)$	(1.66, 0.08)	(3.06, 0.13)	(0.002, 0.067)	(0.06, 0.08)	(0.002, 0.066)	(0.06, 0.07)
$\sigma_b^2$	2	1.04	2.625	1.89	5.42	2.02

Model performance	2004	2005	2006
RMSE	3.34	3.75	5.387
RMSPE	6.44	5.79	6.634
Tr.	0.4	1%	%1

We draw the same general conclusions from model runs in 2005 and 2006 as we did for 2004. Additionally, we should point out that the model's performance remains relatively constant through the years as well as the most of the parameter estimates but there are some parameters that can vary significantly. The interannual variability of some of the model parameters is a result of the observed differences in the meteorological fields and ozone concentration levels. Investigation of the interannual variability (long term ozone trend) is beyond the scope of this study.