Combing Measurements With Deterministic Model Outputs: Predicting Ground-Level Ozone

by

Zhong Liu

B.Sc., The University of Science and Technology of China, 2001
M.Sc., The National University of Singapore, 2003

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

Doctor of Philosophy

in

The Faculty of Graduate Studies

(Statistics)

The University Of British Columbia

Dec, 2007

© Zhong Liu 2007

Abstract

The main topic of this thesis is how to combine model outputs from deterministic models with measurements from monitoring stations for air pollutants or other meteorological variables. We consider two different approaches to address this particular problem.

The first approach is by using the Bayesian Melding (BM) model proposed by Fuentes and Raftery (2005). We successfully implement this model and conduct several simulation studies to examine the performance of this model in different scenarios. We also apply the melding model to the ozone data to show the importance of using the Bayesian melding model to calibrate the model outputs. That is, to adjust the model outputs for the prediction of measurements. Due to the Bayesian framework of the melding model, we can extend it to incorporate other components such as ensemble models, reversible jump MCMC for variable selection.

However, the BM model is purely a spatial model and we generally have to deal with space-time dataset in practice. The deficiency of the BM approach leads us to a second approach, an alternative to the BM model, which is a linear mixed model (different from most linear mixed models, the random

effects being spatially correlated) with temporally and spatially correlated residuals. We assume the spatial and temporal correlation are separable and use an AR process to model the temporal correlation. We also develop a multivariate version of this model.

Both the melding model and linear mixed model are Bayesian hierarchical models, which can better estimate the uncertainties of the estimates and predictions.

Table of Contents

Abstract ii

Table of Contents iv

List of Tables vii

List of Figures xiii

Notations, Abbreviations and Conventions xix

Acknowledgments xxii

Dedication xxiv

1 Introduction 1

2 Spatial Covariance and Melding Model 7

2.1 Definition and Estimation of The Variogram 7

2.2 A Class of Non-stationary Spatial Models 11

2.3 Kriging 17

2.4 Bayesian Melding Model 18

3 Implementation of the Bayesian Melding Model 23

3.1 Bayesian Melding in Stationary Case 23

3.2 Bayesian Melding in Non-stationary Case 30

3.3 MCMC Algorithm 31

3.4 Spatial Prediction 33

3.5 Extension to Ensembles 34

3.6 Reversible Jump MCMC 37

3.7 Bayesian Melding Model Software 40

4 Testing the Bayesian Melding Model 47

4.1 A Single Deterministic Model 47

4.2 Ensemble Studies 58

4.3 Reversible jumps: the stationary case 62

4.4 Bayesian Melding in Non-stationarity Case 64

5 A New Univariate Spatial-Temporal Model 74

5.1 Two-Step Linear Regression 74

5.2 Two-step linear regression with Kriging 79

5.3 A Bayesian Hierarchical Spatial-Temporal Model 79

6 Multivariate Spatial-Temporal Model 93

6.1 Multivariate Spatial-Temporal Model 95

6.2 MCMC Algorithm 98

7 Forecasting and Spatial Prediction 102

7.1 Bayesian Melding Model Prediction 102

7.2 Summary and Conclusions 122

7.3 Spatial-temporal Model Prediction and Forecasting 126

7.4 Multivariate Spatial-temporal Model Prediction and Forecast 136

8 Calibration of Deterministic models 145

9 Application of the Bayesian Ensemble Melding Model . . . 153

9.1 Data Description 153

9.2 Data Analysis 155

10 Dynamic Bayesian Ensemble Melding Model 165

10.1 DeGroot's problem 165

10.2 Dynamic Bayesian Ensemble Melding forecast 167

11 Discussion and Future Work 172

11.1 Bayesian Melding Model 172

11.2 Bayesian Spatial-Temporal Model 175

Bibliography 178

List of Tables

4.1 Estimates of a and b with different numbers of model outputs. 52

4.2 Estimates of $\boldsymbol{\beta}$ with different numbers of model outputs. . . . 53

4.3 Estimates of $\boldsymbol{\theta} = (\sigma, \rho)$ with different numbers of model outputs. 54

4.4 SSPE (Sum of squared prediction errors) of Kriging and the BM model with different number of grid cells. Since the data is simulated, there is no unit associated to the numbers in this table. 66

4.5 The SSPE when the true values of the parameters (a, b, $\boldsymbol{\beta}$, $\boldsymbol{\theta}$) are used. That is, we use the conditional mean of the unavailable measurements based on the available measurements to predict at the 100 unavailable stations. Since the data is simulated, there is no unit associated to the numbers in this table. 67

4.6 Coverage probability of the credible interval for the simulation study in Section 4.1, with 20 monitored sites and up to 50 grid cells to predict 100 un-monitored sites. The first column is the nominal coverage probability of the credible interval. 68

4.7 Estimates are based on 20 monitored sites and 50 sampling points. There are 25 sampling points in close proximity to one another. The averages are computed over 50 dataset. True values are $\sigma = 1.50$ and $\rho = 5.00$. Since this table is for the estimates of parameters σ and ρ, there is no unit associated to the numbers in this table. 68

4.8 Parameter estimates in the simulation study of ensembles of deterministic models in Section 4.2. Columns 3-4 give the averages of posterior means and standard deviations for a_i, b_i, $i = 1, 2$, the calibration parameters of the model output for two deterministic models. $\sigma_{\delta,1}$ and $\sigma_{\delta,2}$ are the variances of two model output error processes. The averages are computed over 15 dataset. Since this table is for the estimates of parameters, there is no unit associated to the numbers in this table. Each of these parameters is a scalar. 69

4.9 SSPE of the BM model and Kriging in the simulation of multi-deterministic models. We have 50 monitored sites and 100 unmonitored ones. There are 20 grid cells each of them having two sampling points inside. Since the data is simulated, there is no unit associated to the numbers in this table. 70

4.10 Estimates in the stationary case using reversible jump BM model. Both the "estimates" and "standard error" are averages obtained from 15 independent dataset. Since this table is for the estimates of parameters, there is no unit associated to the numbers in this table. Each of these parameters is a scalar. 71

4.11 SSPEs with different choices of h in non-stationary BM model. Columns 2-6 give the SSPEs for different ratios of the true bandwidth generating the data over the chosen bandwidth in the non-stationary BM model. Since the data is simulated, there is no unit associated to the numbers in this table. 72

4.12 Estimates of a and b for varying values of h in non-stationary BM model. Columns 2-5 give the averages of posterior means and standard deviations. The averages are computed over the 15 datasets generated in the simulation study. The true values are $a = 5.00$ and $b = 2.50$. Since this table is for the estimates of parameters, there is no unit associated to the numbers in this table. Each of these parameters is a scalar. 73

7.1 Average RMSPE of hourly ozone predictions at 30 stations. Column 1: RMSPE for BM model. Column 2: RMSPE for Kriging using measurements. Column 3: RMSPE for Kriging using model outputs. The unit of the numbers in this table is ppbv. 119

7.2 RMSPE (root mean square prediction error) for predicting daily average of ozone levels at 30 stations. Column 1: days. Column 2: RMSPE for BM model. Column 3: RMSPE for Kriging with measurements. Column 4: RMSPE for Kriging with model outputs. The number with * is the smallest number in each row. The unit of the numbers from column 2 to 4 in this table is ppbv. 121

7.3 RMSPE (root mean square prediction error) for predicting weekly average of ozone levels at 30 stations. Column 1: days. Column 2: RMSPE for BM model. Column 3: RMSPE for Kriging with measurements. Column 4: RMSPE for Kriging with model outputs. The number with * is the smallest number in each row. The unit of the numbers from column 2 to 4 in this table is ppbv. 123

7.4 Posterior means and standard deviations of parameters $a_1, \cdots, a_5$, $c_1, \cdots, c_5$, σ_ϵ^2, λ_ϵ, λ_a, σ_a^2, σ_c^2, λ_a and ρ. Since this table is for the estimates of parameters, there is no unit associated to the numbers in this table. Each of these parameters is a scalar. 140

7.5 The first column gives the indeces of 15 stations where the first 240 hours' measurements are available. The second, third and forth columns are the RMSFE (square root of the mean square forecast error) of the next 240 hours' measurements fore-casted by the Bayesian spatial-temporal model, ad-hoc approach and model outputs. The number followed by $*$ indicates the approach that has the smallest RMSFE in that row. The unit of the numbers from column 2 to 4 in this table is ppbv. 141

7.6 RMSPE at ungauged sites 1-35. Column 2: Bayesian spatial-temporal model; column 3: ad-hoc approach; column 4: model outputs; column 5: Kriging with measurements. The number followed by $*$ indicates the approach that has the smallest RMSPE in each row. The unit of the numbers from column 2 to 5 in this table is ppbv. 142

7.7 RMSPE at ungauged sites 36-63. Column 2: Bayesian spatial-temporal model; column 3: ad-hoc approach; column 4: model outputs; column 5: Kriging with measurements. The number followed by $*$ indicates the approach that has the smallest RMSPE in each row. The unit of the numbers from column 2 to 5 in this table is ppbv. 143

7.8 Posterior mean and standard deviations of the parameters μ_β, ρ and Σ. $\mu_{\beta,1,1}$ and $\mu_{\beta,2,1}$ are the means of intercepts; $\mu_{\beta,1,2}$ and $\mu_{\beta,1,3}$, the mean of coefficients of day- and nighttime model outputs of the daytime measurements; $\mu_{\beta,2,2}$ and $\mu_{\beta,2,3}$, the mean of coefficients of day- and nighttime model outputs on the nighttime measurements. Since this table is for the estimates of parameters, there is no unit associated to the numbers in this table. Each of these parameters is a scalar. 144

7.9 Average of RMSPE and RMSFE for Bayesian multivariate spatial-temporal model (6.1) (columns 2 and 4) and univariate spatial-temporal model (5.5) (columns 3 and 5). The average of RMSPE is over 63 sites and the average of RMSFE is over 15 stations. The unit of the numbers in this table is ppbv. . . 144

8.1 RMSPE of uncalibrated and calibrated model outputs for the prediction of measurements at the 30 stations. column 1: day 1-30; column 2: RMSPE of uncalibrated model output; column 3: RMSPE of BM calibrated model output; column 4: RMSPE of Bayesian spatial-temporal model calibrated model output; The number with a * indicates the "winner" in that row. The unit of the numbers from column 2 to 4 in this table is ppbv. 149

List of Figures

4.1 Locations of 120 (20 monitored, 100 to be predicted) sites and points sampled from up to 50 grid cells. Monitored sites (Δ), sites with responses to be predicted: $+$, 1: the first 2 points from grid cells, 2: 3-10 points, 3: 11-20 points, 4: 21-30 points, 5: 31-50 points. 50

4.2 Spatial predictions for 100 un-monitored sites by using BM model in the case of 50 grid cells. 55

4.3 MCMC Gibbs samplers of the additive and multiplicative calibration parameters, a and b, respectively, in the case of 50 grid cells. 56

4.4 MCMC Gibbs samplers of coefficients $\boldsymbol{\beta}$ in the case of 50 grid cells. 57

4.5 MCMC Gibbs samplers of covariance parameters $\boldsymbol{\theta} = (\sigma, \rho)$ in the case of 50 grid cells. 58

4.6 MCMC Gibbs samplers of error variance σ_δ^2 and σ_e^2 in the case of 50 grid cells. 59

4.7 Locations of 150 sites (50 monitored, 100 to be predicted) and 20 grid cells. Stations: Δ; sites with sites to be predicted: $+$; grid cells: rectangles. Each grid cell has two sampling points in it. The x-axis is the longitude in degrees, y-axis, the latitude in degrees. 60

4.8 The upper plot: the dimension k of the coefficient $\boldsymbol{\beta}$, as a function of the number of MCMC iterations. The lower plot: histogram of the posterior MCMC samples of k. 63

5.1 Histogram of Pearson's correlation coefficients between hourly measurements and model output for the 78 stations. 75

7.1 Scatter plots and Pearson correlations between measurements of station #290470003 and model outputs for cell #1847. The y-axis stands for the model outputs of grid cell #1847 and the x-axis stands for the measurements at station #290470003. The station #290470003 is inside grid cell #1847. Observe that the correlation between measurements and model outputs is 0.7. 105

7.2 The upper panel is the 24 side-by-side hourly box plots of measurements at station #10731005. The lower panel is the 24 side-by-side hourly box plots of model outputs (simulated ozone levels) in grid cell #3529. The station #10731005 lies inside grid cell #3529. 108

7.3 Histograms of correlation over space at each hour. The upper histogram is for all the 24 hours. The lower histogram is for the 8-hour day time. Notice the markedly larger model-to-measurement correlations during the hours 1000-1700. 109

7.4 Locations of 48 available stations, 78 grid cells and 30 unavailable stations. Δ: available stations. Rectangle: grid cells. $+$: unavailable stations to be predicted. The x-axis is the longitude in degrees, y-axis, the latitude in degrees. 111

7.5 Cumulative plots of the relative errors in percentages between the distance using formula 7.1 and formula 7.2. 112

7.6 Hourly RMSPE (root mean square prediction error) difference between Kriging using measurements and BM versus the correlation over space between measurements and model outputs. Points above the horizontal line in the plot represent victories for the BM model over Kriging. 114

7.7 RMSPE difference between Kriging prediction with measurements and the BM prediction versus the correlation over space between measurements and model outputs. Points above the plotted horizontal line represented victory for the BM model. Notice the supremacy of the BM model when correlation exceeds 0.6. 118

7.8 Daily RMSPE difference between Kriging with measurements and the BM model versus the variance of $\hat{b}$ (defined by formula (7.4)) across space. Points above the plotted horizontal line represented victory for the BM model. Notice the supremacy of the BM model when the variance of $\hat{b}$ over space is small. . 119

7.9 Daily RMSPE difference between Kriging with only measurement and the BM model versus the absolute mean difference of $\hat{b}$ (defined by formula (7.4)) between the measured and unmeasured stations. Points above the plotted horizontal line represented victory for the BM model . Notice the supremacy of the BM model when the absolute mean difference of $\hat{b}$ over space is small. 120

7.10 Plots of partial ACF and ACF for the estimated residual $\hat{\epsilon}_t$ at one station. 128

7.11 Plots of the parameter values in the Gibbs sampling iterations. Upper left panel: a_1; upper right panel: rho; lower left panel: σ_ϵ^2; lower right panel: λ_ϵ. 130

7.12 Plot of measurement and the forecast and with 90% credible interval using the Bayesian spatial-temporal model at one of the 15 stations. 131

7.13 Plot of measurement and the spatial prediction with 90% credible interval from the Bayesian spatial-temporal model at one of the 63 ungauged sites. 133

7.14 Plot of distance versus correlation for the normalized residuals from model 5.5. 134

8.1 Measurements versus the uncalibrated and calibrated model output of 30 grid cells on day 2. The solid line shows the measurements inside the grid cells. The dotted and dashed lines are the uncalibrated and calibrated model outputs respectively. 148

8.2 Scatter plots of uncalibrated model outputs versus BM calibrated model outputs. The solid lines have intercept 0 and slope 1. The plots are for days 2, 7, 8 and 26 from the upper left to the lower right corner in left-to-right sequence. 150

8.3 Scatter plots of uncalibrated model outputs versus Bayesian spatial-temporal model calibrated model outputs. The solid lines have intercept 0 and slope 1. The plots are for days 2, 7, 8 and 26 from the upper left to the lower right in left-to-right sequence. 151

8.4 Histograms of posterior means of a_0, a_1, a_2 and b during the 30 days on which the current study focuses on. 152

9.1 Histograms of 102 hours' Pearson's correlation coefficients between measurement and model output over space. Panels 1-5 are the histograms for model output from five different MM5 initial conditions. 156

9.2 Histograms of posterior mean of a. Panels 1-5 are the histograms of a for model output from five different MM5 initial conditions. 161

9.3 Histograms of posterior mean of b. Panels 1-5 are the histograms of b for model output from five different MM5 initial conditions. 162

9.4 RMSPE histograms of predictions by BEM, model averaging with equal weight and Kriging with measurements only. 163

9.5 The x-axis represents the mean correlation coefficients between measurement and model output, i.e. the mean of the five correlation coefficients between the measurement and five model outputs. 164

Notations, Abbreviations and Conventions

We follow the widely used conventions throughout the thesis. Latin upper-case letters, often X, Y, Z, sometimes with subscripts such as s, t, are used for random variables. Bold Latin upper-case letters, often $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$, are used for random vectors. Bold Greek lower-case letters stand for parameter vectors, e.g., $\boldsymbol{\beta}$, $\boldsymbol{\theta}$. For a vector or matrix, the transpose is indicated with a superscript of T.

List of notation and abbreviations:

$\Re$	$(-\infty, \infty)$
$\sim$	distributed as
$\propto$	proportional to
$\approx$	approximate to
$\otimes$	Kronecker product
$\boldsymbol{I}$	identity matrix
ϕ	multivariate normal density function

$\pi(\cdot)$	prior distribution
$X\|Y$	random variable X conditional on random variable Y
$\hat{\alpha}$	Estimate of α
$N(\mu, \sigma^2)$	univariate normal distribution with mean μ and variance σ^2
$\mathrm{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	multivariate normal distribution with mean $\boldsymbol{\mu}$ and varinace matrix $\boldsymbol{\Sigma}$
Φ	probability function of the standard normal distribution
$\pi(\beta)$	prior distribution of parameter β
$i.i.d$	independently and identically distributed
$\mathrm{AR}(p)$	autoregressive of order p
$\mathrm{E}(X)$	expectation of random variable X
$\mathrm{Var}(X)$	variance of random variable X
$\mathrm{Cov}(X, Y)$	covariance of random variables X and Y
$\{Z\}$	a stochastic process
$\boldsymbol{s}$	a location in a geographic region
B	a grid cells of the deterministic model outputs in a geographic region
$\boldsymbol{x}$	also indicate a location in a geographic region in Section 2.2
$\boldsymbol{u}$	also indicate a location in a geographic region in Section 2.2
$\boldsymbol{h}$	a distance vector between two locations in a geographic region
$C(\cdot)$	the covariance function
$\gamma(\cdot)$	the variogram function
σ	variance of a stochastic process also known as the sill parameter
ρ	range parameter of a covariance function $C(\cdot)$ also represent auto-regressive parameter in Chapters 5, 6 and Sections 7.3 and 7.4

t	indicate time in hours in Chapter 5 and Section 7.3, in days in Chapter 6 and Section 7.4
τ	nugget parameter of a covariance function $C(\cdot)$
$K(\cdot)$	the kernel function
ANOVA	analysis of variance
$\{\hat{Z}\}$	measurement process over the space
$\{\tilde{Z}\}$	deterministic model outputs process over the space
$\sim$	"distributed as" or "has its distribution in"
$\approx$	"approximate to"
$f(x)$	density function of random variable $X = x$
MCMC	Markov chain Monte Carlo
sd	standard deviation
ppbv	part par billion by volume
BM	Bayesian melding
BEM	Bayesian ensemble melding
DLM	dynamic linear model
RMSPE	root mean square prediction error
RMSFE	root mean square forecast error
SSPE	sum square prediction error
DBEM	dynamic Bayesian ensemble melding
model output	deterministic model outputs
station	gauged site where measurements are available

Acknowledgments

There are so many people to thank. First of all, I thank my two supervisors, Jim Zidek and Nhu Le, for their academic and financial support. Without their patient coaching and continual encouragement, there is no way I can finish my thesis. Both of them are great mentors and I feel truly honored to work with them.

I want to thank my committee member Paul Gustafson,, for his great feedback on my work. Thank you to Jiahua Chen and Douw Steyn who kindly agreed to be University Examiners at my defense. I also thank to the current and previous office staff, Christine Graham, Rhoda Morgan, Elaine Salameh, Peggy Ng and Viena Tran, for all their help with administrative matters.

I am also indebted to my fellow graduate students (Lawrence McCandless, Jean-Franois Plante, Kenneth Lo, etc) for many helpful talks on my research. Special thanks to Mike Danilov, for his patience to answer my many questions about computing issues.

Outside the department, I am grateful to Fahimah Al-Awadhi who helped me to understand how reversible jump MCMC works. Thank you also to

Prasad Kasibhatla who provide ozone air pollution data and MAQSIP modeling outputs. Finally, a note of thanks to Tilmann Gneiting for very generously pointing the way to the MURI ensemble daatset.

On a personal note, I want to say thank you to my family and friends for their continual love and support. I am so lucky to have my mom who always encouraged me to pursue what I truly wanted in my life even when it made life difficult for her.

Finally, I thank my girl friend, Jiaoyang Jia, for her love and support. She has always had a faith in me even when I lost faith in myself.

Dedication

To my father,

Renxi Liu (1949-2001),

whose love is with me always.

Chapter 1

Introduction

Deterministic models have been widely used in scientific research because they have a number of advantages. First and foremost they, unlike statistical models, do incorporate prior knowledge about the underlying processes. The deterministic models usually are made of a series of differential equations with boundary conditions. Since the boundary conditions and inputs of the deterministic models are adjustable, they enable scenario analysis under hypothetical changes to an existing regime, say as a result of abatement strategies. They enable computer experiments to be run when it is not feasible to make the phenomenon they represent happen in real life due to ethical or other constraints. Such experiments can enable the study of input - output relationships and possibly suggest optimal control settings for use in the real phenomenon they represent. One of the most important features of the deterministic models is that repeated runs yield the same outputs with fixed input. That is why they are called "deterministic". Unlike statistical models, they do attempt to capture the fundamental dynamic processes that govern the phenomenon of interest. The complexity of the equations that describe such dynamics can render analytic solution impossible and force the

use of numerical models, typically with lots of complicated computer code and slow running times.

In this thesis, we are interested in the deterministic models used in climate and environmental studies, one example being the RCM (Regional Climate Model) described in Caya et al. (1995). Another example, the MAQSIP (Multiscale Air Quality Simulation Platform) model is described in Odman and Ingram (1996). The chemical transport model (CTM) called GEOS-CHEM (Goddard Earth Observing System-Chemistry), which models hourly ozone fields, is suggested in the current review of NAAQS (National Ambient Air Quality Standards) as a method for estimating ozone's policy relevant background (PRB) level (Garner et al. (2005)). More details about the various versions of GEOS-CHEM models can be found at http://www-as.harvard.edu/chemistry/trop/geos/. The PRB level, a baseline for the NAAQS, is the ozone concentration that would obtain if there were no anthropogenic sources of ozone in North America. The PRB cannot be measured but has to be modeled because even remote areas in North America is polluted by ozone which drifts from other places. For example, the monthly maximal hourly ozone concentration level in the Yellowstone national park (WY) is over 50 ppbv (parts per billion by volume) in most months from 1998 to 2001 (Figure 3-25a in Garner et al. (2005)). This ozone concentration level is not low even compared with urban areas. In the EPA report Garner et al. (2005), the current deterministic models estimate the PRB level is 15-35 ppbv. Moreover the deterministic models can generate

useful byproducts such as estimates of the fraction of the ozone field from the vertical transport from the stratospheric ozone to the troposphere. From now on, we also refer the deterministic model outputs as model outputs for simplicity.

Deterministic models simulate the ozone air pollution on grid cells with certain resolutions. However, the measurements are obtained at each monitoring station over the space. So we cannot compare the model outputs with the measurements directly. Nevertheless, we can examine whether there is correlation between model outputs and the local measurements. If there is such correlation, we can use statistical models to combine the measurements with model outputs to achieve better prediction. In addition, we can also calibrate deterministic model outputs for predicting the ozone concentration levels

Edwards (1967) and Speed (1983) have started studying the deterministic model outputs by using statistical model. Sacks et al. (1989) approximate the experiment outputs with stochastic processes and choose input values to optimize some functions of the outputs. Poole and Raftery (2000) use a Bayesian model to combine different priors for the input to better evaluate the uncertainties of the model outputs. Ramsay et al. (2007) use non-parametric methods to estimate the parameters of differential equations in a deterministic models. However, the complexity of the deterministic models such as MAQSIP sometimes may stop us working on those differential equations directly. So we simply treat the deterministic models as black boxes. We

only focus on the model outputs without learning details of the differential equations in those deterministic models.

Kasibhatla and Chameides (2000); Hogrefe et al. (2001b,a) study deterministic model outputs in meteorological and environmental science contexts by using traditional scatter plots and least squares analysis of measurements and modeling outputs; moreover they examining the correlation between measurements from monitoring stations and deterministic modeling output. Guttorp and Walden (1987) address these issues within a statistical framework. Of particular interest is the work of Fuentes et al. (2003) and Fuentes and Raftery (2002, 2005), in which the Bayesian melding (BM) model is used to combine measurements and modeling output while respecting their intrinsic differences. The combination highlights the discrepancies and similarities between the values. The BM model can be used to predict the phenomenon of interest and calibrate the deterministic models by using both measurements and modeling output. Note that the terminology "Bayesian melding" was first proposed by Poole and Raftery (2000). To avoid confusion, from now on in this thesis, the Bayesian melding (BM) model specifically refers to the model proposed by Fuentes and Raftery (2005). Another Bayesian model is proposed by Sanso and Guenni (2002) who like Fuentes and Raftery (2005), postulate a "true underlying process" but it is specifically concerned with rainfall and uses a truncated model for the relationship between the true underlying process and the data. Berliner (2003); Wikle and Berliner (2005) also talk about how to use Bayesian models to combine information

from different sources. The BM model is built on a Bayesian framework, which enables us to combine measurements and modeling outputs from more than one deterministic models. This extension of the BM model to include model outputs from more than one deterministic models is of interest because usually an ensemble of deterministic models is used in weather forecasting. Raftery et al. (2005) uses a Bayesian model averaging approach to combine model outputs from deterministic models in different initial conditions to give a more accurate weather forecast. The Bayesian ensemble melding (BEM) model can provide an alternative to that approach.

Although the BM model can be used to combine measurements with model outputs, it was designed as a spatial model which cannot handle the temporal correlation appearing in most environmental or climatic data. The ozone data used by Kasibhatla and Chameides (2000) have the hourly measurements and model outputs of the ozone concentration level in eastern and central USA, which has very strong temporal auto-correlation. It is desirable to find an alternative which are spatial-temporal models. Wikle and Cressie (1999); Wikle and Royle (1999); Wikle et al. (2001); Wikle (2001) has proposed spatial-temporal models based on various approaches such as dynamic kalman filter, kernel-based spectral analysis or Bayesian hierarchical model. Guillas et al. (2006) provides a simple but powerful regression approach which enables us to analyze the data with temporal correlation. We first extend this regression approach to a more mathematically rigorous Bayesian spatial-temporal model. Then we extend this Bayesian spatial-

temporal model further to a multivariate model.

The rest of the thesis is organized as follows. Chapter 2 gives the BM model specifications. Then Chapter 3 presents the MCMC algorithm to fit the BM model. As an byproduct, we develop the R code needed to implement the BM model under different scenarios. We also include a description of the code in Chapter 3. We test the BM model with a comprehensive simulation study in Chapter 4. In the simulation study, we investigate the prediction and calibration performance of the BM model in various scenarios. As an alternative to the BM model, Chapter 5 and Chapter 6 present the Bayesian spatial-temporal models in univariate and multivariate cases. Chapter 7 summarizes all the forecast and spatial prediction results for the BM model and univariate/multivariate Bayesian spatial-temporal models. Chapter 8 specifically discusses how we calibrate the deterministic models by using the BM and Bayesian spatial-temporal models. The data source for the analyses in Chapter 7 and 8 are the ozone data used by Kasibhatla and Chameides (2000). Chapter 9 presents an application for the BEM model of the temperature data used by Raftery et al. (2005). Chapter 10 presents a dynamic BEM model which extends the BEM to a spatial-temporal model. This thesis concludes with Chapter 11 in which we discuss some future research topics in assessing or calibrating deterministic models used in climatic or environmental studies.

Chapter 2

Spatial Covariance and Melding Model

This chapter first presents some background knowledge of the stationary and non-stationary spatial covariance structure. Then it presents the specifications of the BM model proposed by Fuentes and Raftery (2005); Fuentes et al. (2003); Fuentes and Raftery (2002).

2.1 Definition and Estimation of The Variogram

Suppose we have a real-valued random process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in D\}$, which is observed at locations $\{\boldsymbol{s}_i : i = 1, ..., n\}$ over a geographic region $D \subset \Re^d$, d being a positive integer. The random process $\{Z\}$ is defined as second order

stationary if it satisfies for all $\boldsymbol{s}$

$$\begin{aligned}
&\mathrm{E}(Z(\boldsymbol{s}+\boldsymbol{h})-Z(\boldsymbol{s}))=0,\\
&\mathrm{Var}(Z(\boldsymbol{s}+\boldsymbol{h}))=\mathrm{Var}(Z(\boldsymbol{s})),\\
&\mathrm{Cov}(Z(\boldsymbol{s}+\boldsymbol{h}),Z(\boldsymbol{s}))=C(\boldsymbol{h}),
\end{aligned}$$

where $C(\cdot)$ is a covariance function. In other words, the correlation between responses at two locations depends only on their degree of separation. In addition, the first and second order moments of the random process $Z(\boldsymbol{s})$ are the same for all $\boldsymbol{s}$. Furthermore, if $C(\boldsymbol{h}) = C(||\boldsymbol{h}||)$, where $||\boldsymbol{h}||$ is the length of $\boldsymbol{h}$, the covariance function $C(\cdot)$ is called isotropic. Another very important quantity used in spatial statistics, the variogram, is related to the spatial covariance between two locations. We have

$$\mathrm{Var}(Z(\boldsymbol{s}+\boldsymbol{h})-Z(\boldsymbol{s}))=\gamma(\boldsymbol{h}),$$

$\gamma(\boldsymbol{h})$ being known as the variogram.

The classical estimator of the variogram proposed by Matheron (1962) is obtained by using the method of moments. The estimator is

$$\hat{\gamma}(\boldsymbol{h})=\frac{1}{|N(\boldsymbol{h})|}\sum_{N(\boldsymbol{h})}(Z(\boldsymbol{s}_i)-Z(\boldsymbol{s}_j))^2, \tag{2.1}$$

where the sum is over $N(\boldsymbol{h}) = \{(i,j) : \boldsymbol{h}-\boldsymbol{\delta} \leq |\boldsymbol{s}_i - \boldsymbol{s}_j| \leq \boldsymbol{h}\}$ and $|N(\boldsymbol{h})|$

is the number of distinct elements in $N(\boldsymbol{h})$. Although unbiased, this classical estimator is highly affected by outliers due to the squared term in the summation. Hawkins and Cressie (1984) presented a more robust estimator

$$\bar{\gamma}(\boldsymbol{h}) = \frac{\{\frac{1}{|N(\boldsymbol{h})|}\sum_{N(\boldsymbol{h})}|Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)|^{1/2}\}^4}{0.457 + \frac{0.494}{|N(\boldsymbol{h})|}}. \tag{2.2}$$

From our definition of the variogram, it is clear that $\gamma(\mathbf{0}) = 0$. However, more generally $\gamma(\boldsymbol{h}) \to \tau > 0$ maybe allowed as $\boldsymbol{h} \to \mathbf{0}$, in which case τ is the called nugget effect by Matheron (1962). One of the sources of the nugget effect is measurement error and commonly measurements are assumed to be the ground truth $Z(\boldsymbol{s})$ plus some measurement error. The book by Cressie (1993) gives more details about the nugget effect.

The covariance function only needs to ensure the covariance matrix it generates is positive definite and symmetric. In the book Stein (1999), Bochner's theorem specifies that the positive definite covariance functions are those which are Fourier transforms of non-negative Borel measures. One commonly used covariance function proposed by Matern (1960) has the following form:

$$C_{\boldsymbol{\theta}}(d) = \frac{\sigma}{2^{\nu-1}\Gamma(\nu)}(2\nu^{1/2}|d|/\rho)^{\nu}K_{\nu}(2\nu^{1/2}|d|/\rho),$$

d being the distance between any two locations. For convenience, we let $\boldsymbol{\theta} = (\sigma, \rho, \nu)$. Here σ, the *sill* parameter, represents the variance of random process $Z(\boldsymbol{s})$ while ρ, the *range* parameter, determines how fast the correlation decreases when distance d increases and ν, the *smoothing* parameter

controls the smoothness of the covariance function. K_ν is the modified Bessel function of type III as described by Abramowitz and Stegun (1972). Matern covariance functions have the advantage of flexibility. Certain choices of the smoothing parameter ν reduce it to simple well known covariance functions. For example, $\nu_s = 1/2$ gives the so-called exponential covariance function,

$$C_\theta(d) = \begin{cases} \sigma \exp(-|d|/\rho) & \text{if } |d| > 0; \\ \sigma + \tau & \text{if } |d| = 0. \end{cases} \tag{2.3}$$

The *nugget* parameter is τ. It is obvious that the covariance functions defined above are only suitable for stationary random fields because we assume the covariance $C_\theta(d)$ is only a function of the distance d. However, in many cases, it is not reasonable to assume second order stationarity. In recent years, the non-stationarity problem has received a substantial amount of attention. One of the earliest and most important papers is due to Sampson and Guttorp (1992), who obtain a semi- parametric estimate of the non-stationary spatial covariance function by transforming the original geographical map into another, the deformed plane. However that deformation method relies on repeated measurements to give an empirical estimate of the variogram, a serious limitation in the context of geostatistics, for example, where typically just one realization of a space - time process is available. Haas (1995, 1998) circumvents that problem with a moving windows method as does Higdon et al. (1999) with a convolution approach although these methods may require data from a large number of monitoring sites to be effective. Other

non-stationary papers include Paciorek and Schervish (2004); Nychka et al. (2002); Paciorek and Schervish (2006). The paper of Gelfand et al. (2004) considers the non-stationarity problem for a multivariate case. Fuentes and Smith (2001) propose a class of non-stationary spatial models used in Fuentes et al. (2003), Fuentes and Raftery (2002), Fuentes and Raftery (2005). This approach's appeal derives not only from its circumvention of the need for replicate measurements but as well, from its simple intuitive, easy to understand formulation as well as the ease with which it can be implemented. More details about the convolution approach are presented in the next section.

2.2 A Class of Non-stationary Spatial Models

Calder and Cressie (2006) gives an excellent summary of the convolution-based spatial models which become more popular recently because of its flexibility. Yaglom (1987) stated that a stationary spatial process $\{Z(\boldsymbol{s})\}$ can be written as

$$Z(\boldsymbol{s}) = \int_{\Re^d} K(\boldsymbol{s}, \boldsymbol{u}) W(d\boldsymbol{u}), \tag{2.4}$$

where $K(\boldsymbol{s}, \boldsymbol{u})$ is a kernel function satisfying $\int K^2(\boldsymbol{s}, \boldsymbol{u}) < M < \infty$ and $W(\cdot)$ is a Brownian motion. The covariance function of the process $\{Z(\boldsymbol{s})\}$ is

$$C(Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2)) = \int_{\Re^d} K(\boldsymbol{s}_1, \boldsymbol{u}) K(\boldsymbol{s}_2, \boldsymbol{u}) d\boldsymbol{u}.$$

To represent a non-stationary spatial process, we can modify 2.4 by either assuming the kernel function $K(\boldsymbol{s}, \boldsymbol{u})$ or the Brownian motion process $W(\cdot)$ depending on location $\boldsymbol{s}$. Higdon et al. (1999) use the convolution approach with a spatially dependent kernel function to represent non-stationary process. That is,

$$Z(\boldsymbol{s}) = \int_{\Re^d} K_s(\boldsymbol{s}, \boldsymbol{u}) W(d\boldsymbol{u}), \tag{2.5}$$

whose covariance function is

$$C(Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2)) = \int_{\Re^d} K_{\boldsymbol{s}_1}(\boldsymbol{s}_1, \boldsymbol{u}) K_{\boldsymbol{s}_2}(\boldsymbol{s}_2, \boldsymbol{u}) d\boldsymbol{u}.$$

Higdon et al. (1999) use Gaussian kernel function so that the resulting non-stationary process has a Gaussian covariance function.

Departing from Higdon et al. (1999), Fuentes and Smith (2001) assume the kernel function is constant and convolves stationary process as opposed to Brownian motion process. This section shows how we can use the approach of Fuentes and Smith (2001) to model the non-stationary process in detail. This approach assumes the existence at a number of locations of unobservable, independent, stationary random processes that need not have any physical meaning. Their existence only helps us to construct the observable non-stationary process so they do not need to have any physical meaning. The observed non-stationary process is represented as a weighted average or convolution of these latent processes. Each latent random process has its own covariance parameters which vary from one latent process (location) to

another. The weight attached to each of the stationary processes in the representation of the observed non-stationary process depends on its location. So the covariance of the process between any two locations depends not only on their distance d but also on their locations. The rest of this section gives a detailed account of the non-stationary model of Fuentes and Smith (2001). We include it to enable us to introduce the detailed MCMC algorithm used to fit this model in Section 3.2.

Suppose the latent stationary processes are $Z_i, i = 1, \cdots, K$, with $\mathrm{Cov}(Z_i(\boldsymbol{s}), Z_j(\boldsymbol{s})) = 0$ for $i \neq j$. Then the observed process $Z(\boldsymbol{s})$ can be expressed as

$$Z(\boldsymbol{s}) = \sum_{i=1}^{K} Z_i(\boldsymbol{s}) w_i(\boldsymbol{s}), \ \sum_{i=1}^{K} w_i = 1$$

in which w_i is the weight attached to stationary process Z_i. The above equation can be extended to an integral by replacing the weight with a kernel function as follows:

$$Z(\boldsymbol{x}) = \int_D K(\boldsymbol{x} - \boldsymbol{s}) Z_{\boldsymbol{\theta}(\boldsymbol{s})}(\boldsymbol{x}) d\boldsymbol{s},$$

where $K(\boldsymbol{X} - \boldsymbol{s})$ is a kernel function. So, the weight of latent process $Z_{\boldsymbol{\theta}(\boldsymbol{s})}$ depends on the location difference vector between $\boldsymbol{x}$ and $\boldsymbol{s}$. The spatial covariance parameter vector of the latent stationary process $\{Z_{\boldsymbol{\theta}(\boldsymbol{s})}\}$ is $\boldsymbol{\theta}(\boldsymbol{s})$ depending on its center location $\boldsymbol{s}$. That is why these latent stationary processes are called "locally stationary".

Since the covariance between $Z_{\theta(\boldsymbol{s})}(\boldsymbol{x}_1)$ and $Z_{\theta(\boldsymbol{s})}(\boldsymbol{x}_2)$ is

$$\text{Cov}\left(Z_{\theta(\boldsymbol{s})}(\boldsymbol{x}_1), Z_{\theta(\boldsymbol{s})}(\boldsymbol{x}_2)\right) = C_{\theta(\boldsymbol{s})}(\boldsymbol{x}_1 - \boldsymbol{x}_2),$$

the covariance between $Z(\boldsymbol{x}_1)$ and $Z(\boldsymbol{x}_2)$ in the non-stationary process Z can be expressed as a convolution of the local covariance $C_{\theta(\boldsymbol{s})}(\boldsymbol{x}_1 - \boldsymbol{x}_2)$:

$$C(Z(\boldsymbol{x}_1), Z(\boldsymbol{x}_2)) = \int_D K(\boldsymbol{x}_1 - \boldsymbol{s})K(\boldsymbol{x}_2 - \boldsymbol{s})C_{\theta(\boldsymbol{s})}(\boldsymbol{x}_1 - \boldsymbol{x}_2)d\boldsymbol{s}. \tag{2.6}$$

The process $\{Z(\boldsymbol{x})\}$ is non-stationary because the covariance between $Z(\boldsymbol{x}_1)$ and $Z(\boldsymbol{x}_2)$ depends on the locations x_1 and x_2.

The kernel function can be any of the form $K(u) = \frac{1}{h^2}K_0(\frac{u}{h})$, K_0 being any non-negative function with integral 1. Fuentes and Raftery (2005) use the following kernel function

$$K_0(\boldsymbol{u}) = \frac{3}{4}(1 - u_1^2)_+\frac{3}{4}(1 - u_2^2)_+,$$

with $\boldsymbol{u} = (u_1, u_2)$ and in general, $a_+ = \max\{0, a\}$ for any scalar valued quantity a. This kernel function is chosen because it is easy to compute and it has a compact support. However, other kernel functions can be used as well. The choice of kernel function is not as important as the bandwidth parameter h, which can be any positive scalar valued quantity subject to certain restrictions explained later in this section. Also in Section 4.4, a simulation is carried out to examine the effect of different choices of the

bandwidth h on prediction and parameter estimation.

Our choice of the kernel function implies that for a given pair of locations $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, only the local stationary process whose center location $\boldsymbol{s}$ is within circles with origins $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and radius $\boldsymbol{h}$ will have an effect on their covariance.

In the paper by Fuentes and Smith (2001), the integral (2.6) is approximated by an average. First, the center locations $\boldsymbol{s}_m$, $m = 1, \cdots, M$ of those latent stationary processes are chosen as points on a regular grid over the map. The kernel integral (2.6) is then replaced by

$$C_M(\boldsymbol{x}_1, \boldsymbol{x}_2; \boldsymbol{\theta}) = M^{-1} \sum_{m=1}^{M} K(\boldsymbol{x}_1 - \boldsymbol{s}_m) K(\boldsymbol{x}_2 - \boldsymbol{s}_m) C_{\boldsymbol{\theta}(\boldsymbol{s}_m)}(\boldsymbol{x}_1 - \boldsymbol{x}_2). \quad (2.7)$$

In the non-stationary model, the parameter $\boldsymbol{\theta}$ of the latent stationary processes is a function of its center location $\boldsymbol{s}$. This function can be smooth but instead Fuentes and Smith (2001) assume an additive "ANOVA" (analysis of variance) type model for $\boldsymbol{\theta}$ in terms of center locations $\boldsymbol{s}$. The stationary points are points of a regular grid over the map. For example, for the exponential covariance function, with just two parameters, the *sill* σ and the

range ρ, the "ANOVA" forms would be:

$$
\begin{aligned}
\sigma_{i,j} &= \alpha + r_i + c_j + \epsilon_{i,j} \quad i = 1, \cdots, n_1 \\
\rho_{i,j} &= \alpha^{'} + r_i^{'} + c_j^{'} + \epsilon_{i,j}^{'} \quad j = 1, \cdots, n_2 \\
\boldsymbol{\mu}_\sigma &= (\alpha, r_1, \cdots, r_{n_1}, c_1, \cdots, c_{n_2}) \\
\boldsymbol{\mu}_\rho &= (\alpha^{'}, r_1^{'}, \cdots, r_{n_1}^{'}, c_1^{'}, \cdots, c_{n_2}^{'}) \\
\epsilon_{i,j} &\sim \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Sigma}(\tau_\sigma, \eta_\sigma)) \\
\epsilon_{i,j}^{'} &\sim \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Sigma}(\tau_\rho, \eta_\rho)).
\end{aligned}
\tag{2.8}
$$

In the above model, n_1 and n_2 are the numbers of points in the horizontal and vertical directions respectively. As well, r_i and c_j are main effects of center location's longitude and latitude on σ. The main effects of the center location's longitude and latitude on ρ are $r_i^{'}$ and $c_j^{'}$. In statistics,a design matrix is a matrix that is used in certain statistical models such as ANOVA model. It contains indicator variables (ones and zeros) that indicates the group membership. The design matrix for *sill* σ has a dimension of $(n_1 \times n_2) \times (1 + n_1 + n_2)$. The elements in the first column of the matrix are all 1. For columns $k \in [2, n_1 + 1]$, only the elements from rows $l \in [(k-2) \times n_1 + 1, (k-1) \times n_1]$ are 1. For columns $k \in [n_1 + 2, n_1 + n_2]$, only the elements from rows $l \in [(k - n_1 - 1) \times n_1 + 1, (k - n_1) \times n_1]$ are 1. All the other elements in the design matrix are 0.

In this hierarchical model, σ and ρ might well have normal prior distributions. Then the hyper-parameters would be $r_i, r_i^{'}, c_j, c_j^{'}$ and $\tau_\sigma, \eta_\sigma, \tau_\rho, \eta_\rho$.

This "ANOVA" model is flexible because it does not assume a parametric function of σ and ρ over the space. However, the normality assumption of σ and ρ could be a problem because these two parameters should always be positive. An alternative is to model them with log-normal distributions which ensure σ's and ρ's are positive. Section 3.2 explores an algorithm for fitting this non-stationary model.

2.3 Kriging

Kriging is a celebrated geostatistical approach to spatial interpolation in the point-referenced data setting, a name given to it by Matheron (1963) in honor of D.G. Krige, a South African mining engineer. Kriging, the best linear unbiased predictor (BLUP), weights the available observations in accordance with the distance between the locations where they are made and that of the response to be interpolated. These BLUP weights are obtained by minimizing the variance of the interpolation error assuming the spatial covariance is stationary and known. Given measurements of a random field $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), ..., Y(\boldsymbol{s}_n))^T$ at locations $(\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n)$, the question is how to predict the random variable Y at a new site $\boldsymbol{s}_0$. The simplest Kriging approach, ordinary Kriging, assumes $E(\boldsymbol{Y}_{\boldsymbol{s}_0}) = E(\boldsymbol{Y}_{\boldsymbol{s}_i}) = \mu$, $i = 1, \cdots, n$. The interpolated value at $\boldsymbol{s}_0$ is

$$Y^*(\boldsymbol{s}_0) = \sum_{i=1}^{n} w_i Y(\boldsymbol{s}_i),$$

w_i being the weight of $Y(\boldsymbol{s}_i)$ subject to the constraint $\sum_{i=1}^{n} w_i = 1$. Obviously, $Y^*(\boldsymbol{s}_0)$ is an unbiased estimator of $Y(\boldsymbol{s}_0)$. So, to minimize the MSE (mean square error) of $Y^*(\boldsymbol{s}_0)$, we only need to minimize its variance,

$$\begin{aligned}
\sigma_E^2 &= E\left[(Y^*(\boldsymbol{s}_0) - Y(\boldsymbol{s}_0))^2\right] \\
&= -\gamma(\boldsymbol{s}_0 - \boldsymbol{s}_0) - \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) + 2\sum_{i=1}^{n} w_i \gamma(\boldsymbol{s}_i - \boldsymbol{s}_0).
\end{aligned}$$

Besides ordinary Kriging, universal Kriging assumes the mean of the process depends on other variables such as the coordinates. Co-Kriging is a multivariate extension of Kriging. In practice, the spatial covariance function is estimated empirically by (2.1) or (2.2). Its simplicity and ease of computation has made Kriging very flexible and hence a valuable tool in situations where it is applicable. The Kriging method is available in the R package "geoR" developed by Ribeiro and Diggle (2001), that we use to implement the method in our simulation studies and data analyses later in this thesis.

2.4 Bayesian Melding Model

Although Kriging is easy to implement, Kriging relies on a known variogram to compute its weights, those estimated parameters as fixed. So the true uncertainty in interpolation is underestimated. The simulation study in Section 4.1 demonstrates that disadvantage.

More importantly, grid cell data from a deterministic model will be on

coarser scales of resolution than the micro scale on which measurements are made. For example, the deterministic AQM studied in this thesis output its hourly ozone concentration data at a resolution 6×6 km^2. The mismatch of scales leads to a need to re-calibrate the simulated data when combining it with the data for interpolation. While this could be done in an ad hoc fashion with Kriging, it is not designed to deal with that issue in a fundamental way.

In contrast to Kriging, the BM model, developed and studied by Fuentes and Raftery (2005), Fuentes et al. (2003), as well as Fuentes and Raftery (2002), is designed to do that. It combines measurements and deterministic model outputs in a Bayesian framework that enables calibration of deterministic model outputs and spatial prediction including interpolation.

BM model links processes with responses on mismatched scales through an underlying true process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \Re^D\}$ called the "truth", D being the dimension of the domain. If the location $\boldsymbol{s}$ only has longitude and latitude then $D = 2$, while if the location also has altitude then $D = 3$. The underlying true (latent) process, being unobservable, must itself be estimated. Denote the measurement process by $\{\hat{Z}(\boldsymbol{s}) : \boldsymbol{s} \in \Re^D\}$, and the deterministic model output process by $\{\tilde{Z}(B)\}$, B being the grid cell. To match $Z(\boldsymbol{s})$, we also hypothesize the existence of deterministic model output process $\{\tilde{Z}(\boldsymbol{s}) : \boldsymbol{s} \in \Re^D\}$ based on locations $\boldsymbol{s}$. Of course, the purely conceptual process $\{\tilde{Z}(\boldsymbol{s})\}$ does not actually exist. Its purpose: to link the model output with the truth at the micro- scale, thereby enabling its representation as an integral of $\{\tilde{Z}(\boldsymbol{s})\}$. The truth thus serves as the common basis for both

processes $\{\hat{Z}(\boldsymbol{s})\}$ and $\{\tilde{Z}(\boldsymbol{s})\}$.

The Bayesian model has the following mathematical form:

$$
\begin{aligned}
\hat{Z}(\boldsymbol{s}) &= Z(\boldsymbol{s}) + e(\boldsymbol{s}) \\
Z(\boldsymbol{s}) &= \mu(\boldsymbol{s}) + \epsilon(\boldsymbol{s}) \\
Z(B) &= \frac{1}{|B|}\int_B Z(\boldsymbol{s})d\boldsymbol{s} \\
\tilde{Z}(\boldsymbol{s}) &= a(\boldsymbol{s}) + b(\boldsymbol{s})Z(\boldsymbol{s}) + \delta(\boldsymbol{s}) \\
\tilde{Z}(B) &= \frac{1}{|B|}\int_B a(\boldsymbol{s})d\boldsymbol{s} + \frac{1}{|B|}\int_B b(\boldsymbol{s})Z(\boldsymbol{s})d\boldsymbol{s} + \frac{1}{|B|}\int_B \delta(\boldsymbol{s})d\boldsymbol{s} \\
\mu(\boldsymbol{s}) &= X(\boldsymbol{s})\boldsymbol{\beta}.
\end{aligned} \tag{2.9}
$$

In the above model, the measurements error and model output error are independent of each other. The measurement errors, $e(\boldsymbol{s})$, are independent and identically distributed, having a normal distribution $N(0, \sigma_e^2)$. The model output errors, $\delta(\boldsymbol{s})$, are independent and identically distributed with a normal distribution $N(0, \sigma_\delta^2)$. The spatially correlated residuals, $\epsilon(\boldsymbol{s})$, have zero mean and covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the covariance parameter vector. The number of locations is n. $Z(B)$ and $\tilde{Z}(B)$ are integrals of $Z(\boldsymbol{s})$ and $\tilde{Z}(\boldsymbol{s})$ over grid cell B. We only observe realizations of process $\hat{Z}(\boldsymbol{s})$ and $\tilde{Z}(B)$ at measured stations and grid cells for model outputs.

The mean of the true underlying process is $\mu(\boldsymbol{s}) = X(\boldsymbol{s})\boldsymbol{\beta}$, where $X(\boldsymbol{s})$ is a function of the coordinates at location $\boldsymbol{s}$ and $\boldsymbol{\beta}$ is the corresponding coefficient vector. We use polynomial function although other functions can also be used. The polynomial function can have degree such as 1, 2 up to 3 as

necessary. The covariance matrix $\Sigma(\boldsymbol{\theta})$ is constructed with some covariance functions having a parameters vector $\boldsymbol{\theta}$. (See the previous section for a detailed discussion of covariance functions in stationary and non-stationary cases.) Throughout this thesis, we use the exponential covariance function (2.3). However, the software we have developed and provided can use either the Matern, Gaussian or exponential covariance function. In BM model (2.9), measurement $\hat{Z}(\boldsymbol{s})$ is modeled as the truth $Z(\boldsymbol{s})$ plus measurement error $e(\boldsymbol{s})$ and model output $\tilde{Z}(\boldsymbol{s})$, as $Z(\boldsymbol{s})$ times a multiplicative calibration parameter $b(\boldsymbol{s})$ plus additive calibration parameter $a(\boldsymbol{s})$ and random error $\delta(\boldsymbol{s})$.

In general, we can assume that the calibration parameters are functions of the location of $\boldsymbol{s}$, namely $b(\boldsymbol{s})$ and $a(\boldsymbol{s})$ in the model specification, to take into account their variability with respect to locations. For simplicity, we assume a and b are constants throughout the rest of this chapter for derivation and in the simulation study. But later in the data analysis, we assume a is a function of the coordinates of the location $\boldsymbol{s}$ and b still remains constant. The reason to keep b constant is that a varies over space much more than b.

Fuentes and Raftery (2005) suggest and we use a Monte Carlo method to approximate the integrals in model (2.9). For example, we could sample L points $\boldsymbol{s}_{1,B}, \cdots, \boldsymbol{s}_{L,B}$ within grid cell B. From now on, we call these points "sampling points" to distinguish them from stations. Then, $Z(B)$ can be approximated by the average of the values of process evaluated at the sampling

points within B:

$$Z(B) \approx \frac{1}{L} \sum_{j=1}^{L} Z(\boldsymbol{s}_{j,B}). \tag{2.10}$$

Chapter 3

Implementation of the Bayesian Melding Model

This chapter presents the details of how to fit the BM model (2.9) by using the Gibbs sampling algorithm proposed by Gelfand and Smith (1990). First, we consider the BM model with stationary spatial covariance structure, then we show how to fit the model with non-stationary spatial covariance structure. We also include an extension to the model (2.9) to combine measurements with model outputs from an ensemble of deterministic models. This chapter also presents how to incorporate reversible jump MCMC into the BM model. Finally, we give a description of the R package we have developed to implement the BM model in different scenarios.

3.1 Bayesian Melding in Stationary Case

The Bayesian paradigm primarily seeks the posterior distribution of all the unknowns given the data as a description of their uncertainty. These unknowns include the random error variances $\sigma_e^2, \sigma_\delta^2$, coefficient vector $\boldsymbol{\beta}$, the

true underlying process $\boldsymbol{Z}$, calibration parameters a, b and covariance parameter vector $\boldsymbol{\theta}$. We use vector $\boldsymbol{Z}$ to stand for realizations of the true underlying process at the stations and sampling points within grid cells. If we have n stations and m grid cells, the dimension of $\boldsymbol{Z}$ is $n + m \times L$. The dimensions of the measurements $\hat{\boldsymbol{Z}}$ and model outputs $\tilde{\boldsymbol{Z}}$ are n and m respectively. Let $H = \{\boldsymbol{X}, \hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}}\}$ represent all the data, where $\boldsymbol{X}$ is the covariate matrix in model (2.9).

By using the above notation, the joint distribution of all the unknowns and available data can be decomposed as follows:

$$\begin{aligned}
& p(\hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}}, \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\theta}, a, b, \sigma_e^2, \sigma_\delta^2) \\
= \; & p(\hat{\boldsymbol{Z}}|\boldsymbol{Z}, \sigma_e^2) p(\tilde{\boldsymbol{Z}}|\boldsymbol{Z}, a, b, \sigma_\delta^2) p(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{\theta}) p(\sigma_e^2, \sigma_\delta^2, \boldsymbol{\beta}, \boldsymbol{\theta}) \\
= \; & \Phi_{\boldsymbol{\Sigma}_1}(\hat{\boldsymbol{Z}} - A_0 \boldsymbol{Z}) \Phi_{\boldsymbol{\Sigma}_2}(\tilde{\boldsymbol{Z}} - a - b A_1 \boldsymbol{Z}) \Phi_{\boldsymbol{\Sigma}_3}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta}) p(\sigma_e^2) p(\sigma_\delta^2) p(\boldsymbol{\beta}) p(\boldsymbol{\theta})
\end{aligned} \tag{3.1}$$

Note that given $\boldsymbol{Z}$, $\hat{\boldsymbol{Z}}$ and $\tilde{\boldsymbol{Z}}$ are independent. In (3.1), $\Phi_{\boldsymbol{\Sigma}}(\boldsymbol{\mu})$ stands for the multivariate normal density with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We take the components of $(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_e^2, \sigma_\sigma^2)$ to have independent prior distributions, that is, $p(\sigma_e^2, \sigma_\sigma^2, \boldsymbol{\beta}, \boldsymbol{\theta}) = p(\sigma_e^2) p(\sigma_\sigma^2) p(\boldsymbol{\beta}) p(\boldsymbol{\theta})$. The matrices A_0 and A_1 are

$$A_0 = \begin{pmatrix} 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & 0 & \dots & 0 \end{pmatrix}_{n \times (mL+n)} ;$$

and

$$A_1 = \begin{pmatrix} 0 & \dots & 0 & \frac{1}{L} & \dots & \frac{1}{L} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & \frac{1}{L} & \dots & \frac{1}{L} \end{pmatrix}_{m\times(mL+n)} .$$

In matrix A_0 the first n columns form an identity matrix and all other elements are all zero. In matrix A_1, the elements in row i are $\frac{1}{L}$ from column $n+1+(i-1)\times L$ to $n+1+i\times L$ and all other elements are all zero. We get (3.1) by using the approximation (2.10). $\boldsymbol{\Sigma}_1 = \sigma_e^2 \boldsymbol{I}$ is the covariance matrix of measurement error vector $\boldsymbol{e} = [e(\boldsymbol{s}_1), \cdots, e(\boldsymbol{s}_n)]^{\mathrm{T}}$, $\boldsymbol{\Sigma}_2 = \sigma_\delta^2 \boldsymbol{I}$, the covariance matrix of $\boldsymbol{\delta} = [\delta(B_1), \cdots, \delta(B_n)]^{\mathrm{T}}$ and $\boldsymbol{\Sigma}_3 = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, the covariance matrix of $\boldsymbol{Z}$ while $\boldsymbol{I}$ is the identity matrix.

The density of the joint conditional distribution for $(\hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}}, \boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{\theta}, a, b, \sigma_e^2, \sigma_\delta^2)$ is

$$\begin{aligned}
& p(\hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}}, \boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{\theta}, a, b, \sigma_e^2, \sigma_\delta^2) \\
\propto\ & \exp\left\{-\frac{1}{2}\left[\left(\hat{\boldsymbol{Z}} - A_0\boldsymbol{Z}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_1^{-1}\left(\hat{\boldsymbol{Z}} - A_0\boldsymbol{Z}\right)\right.\right. \\
+\ & \left(\tilde{\boldsymbol{Z}} - a - bA_1\boldsymbol{Z}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_2^{-1}\left(\tilde{\boldsymbol{Z}} - a - bA_1\boldsymbol{Z}\right) \\
+\ & \left.\left.(\boldsymbol{Z} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}_3^{-1}(\boldsymbol{Z} - \boldsymbol{\mu})\right]\right\} \\
=\ & \exp\left\{-\frac{1}{2}\left[-\boldsymbol{Z}^{\mathrm{T}}\left(A_0^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}\hat{\boldsymbol{Z}} + bA_1^{\mathrm{T}}\boldsymbol{\Sigma}_2^{-1}(\tilde{\boldsymbol{Z}} - a) + \boldsymbol{\Sigma}_3^{-1}\boldsymbol{\mu}\right)\right.\right. \\
-\ & \left(\hat{\boldsymbol{Z}}^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}A_0 + b(\tilde{\boldsymbol{Z}} - a)\boldsymbol{\Sigma}_2^{-1}A_1 + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\Sigma}_3^{-1}\right)\boldsymbol{Z} \\
+\ & \left.\left.\boldsymbol{Z}^{\mathrm{T}}\left(A_0^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}A_0 + b^2A_1^{\mathrm{T}}\boldsymbol{\Sigma}_2^{-1}A_1 + \boldsymbol{\Sigma}_3^{-1}\right)\boldsymbol{Z}\right]\right\} + C, \qquad (3.2)
\end{aligned}$$

where $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$ is the mean vector of $\boldsymbol{Z}$ and C has the other terms without $\boldsymbol{Z}$.

As is well known, the normal prior distribution is conjugate to the normal sampling distribution. The full conditional distribution of $\boldsymbol{Z}$ conditional on all the other unknown parameters and available data is also a normal with mean vector $\tilde{\boldsymbol{\mu}}$ and variance matrix $\tilde{\boldsymbol{\Sigma}}$. So, the full conditional distribution of $\boldsymbol{Z}$ must be

$$p\left(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_e^2, \sigma_\delta^2, a, b, H\right) \propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{Z}^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Z} - \tilde{\boldsymbol{\mu}}^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Z}\right]\right\}. \tag{3.3}$$

But what are $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Sigma}}$? By matching the terms containing $\boldsymbol{Z}$ and $\boldsymbol{Z}^{\mathrm{T}}$ in (3.3) and (3.2), we identify the required $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Sigma}}$ for the full conditional distribution of $\boldsymbol{Z}$ as the following:

$$\begin{aligned}
&\boldsymbol{Z}|(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_e^2, \sigma_\delta^2, a, b, H) \sim \mathrm{MVN}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}});\\
&\tilde{\boldsymbol{\Sigma}}^{-1} = (A_0^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}A_0 + b^2A_1^{\mathrm{T}}\boldsymbol{\Sigma}_2^{-1}A_1 + \boldsymbol{\Sigma}_3^{-1});\\
&\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\Sigma}}\left(A_0^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}\hat{\boldsymbol{Z}} + bA_1^{\mathrm{T}}\boldsymbol{\Sigma}_2^{-1}(\tilde{\boldsymbol{Z}} - a) + \boldsymbol{\Sigma}_3^{-1}\boldsymbol{\mu}\right).
\end{aligned} \tag{3.4}$$

If the prior for $\boldsymbol{\beta}$ were $p(\boldsymbol{\beta}) \sim \mathrm{MVN}(\boldsymbol{\beta}_0, \boldsymbol{F})$, then we could claim the full conditional distribution of $\boldsymbol{\beta}$ would be

$$\boldsymbol{\beta}|(\boldsymbol{\theta}, \boldsymbol{Z}, \text{other parameters}) \sim \mathrm{NVN}(\boldsymbol{B}\boldsymbol{b}, \boldsymbol{B}), \tag{3.5}$$

where $\boldsymbol{B}^{-1} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}_3^{-1}\boldsymbol{X} + \boldsymbol{F}^{-1}$ and $\boldsymbol{b} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}_3^{-1}Z + \boldsymbol{F}^{-1}\boldsymbol{\beta}_0$. The proof of

that claim now follows.

Given Σ_3^{-1} and $\boldsymbol{Z}$, $\boldsymbol{\beta}$ is independent of other parameters. So the full conditional distribution of $\boldsymbol{\beta}$ is

$$\begin{aligned} & p(\boldsymbol{\beta}|\boldsymbol{Z}, \Sigma_3) \propto p(\boldsymbol{Z}|\boldsymbol{\beta}, \Sigma_3^{-1})p(\boldsymbol{\beta}) \\ \propto \quad & \exp\{-\frac{1}{2}\left[(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta})^{\mathrm{T}}\Sigma_3^{-1}(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta}) + (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^{\mathrm{T}}\boldsymbol{F}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right]\}. \end{aligned}$$

We can find the mean and variance of $p(\boldsymbol{\beta}|\boldsymbol{Z}, \Sigma(\boldsymbol{\theta}))$ by an approach of Lindley and Smith (1972); Smith (1973).

For the exponential covariance function, the two components σ and ρ of $\boldsymbol{\theta}$ are independent with inverse gamma and gamma prior distributions respectively. The inverse gamma distribution has the density function,

$$p(x; \alpha, \gamma) = \frac{1}{\Gamma(\alpha)\gamma^{\alpha}} \frac{1}{x^{\alpha+1}} e^{-1/(\gamma x)},$$

$\alpha > 0$ and $\gamma > 0$ being the shape and scale parameters, respectively.

The full conditional posterior density function of $\boldsymbol{\theta}$, given the other variables and the data is

$$p(\boldsymbol{\theta}|\boldsymbol{\beta}, \boldsymbol{Z}) \propto p(\boldsymbol{\theta})|\Sigma_3|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta})^{\mathrm{T}}\Sigma_3^{-1}(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta})\right], \qquad (3.6)$$

where $p(\boldsymbol{\theta})$ is the prior density for $\boldsymbol{\theta}$. In general, we have $\Sigma_3 = \sigma g(\rho, \boldsymbol{C})$, where $\boldsymbol{C}$ is the Euclidean distance matrix between the stations and sampling points and $g(\cdot)$ is the spatial correlation function. Then the full conditional

distribution of $\boldsymbol{\theta}$ can be written as

$$p(\boldsymbol{\theta}|\boldsymbol{\beta}, \boldsymbol{Z}) \propto p(\sigma)p(\rho)|\sigma g(\rho, \boldsymbol{C})|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{Z} - \boldsymbol{X\beta})^{\mathrm{T}}(\sigma g(\rho, \boldsymbol{C}))^{-1}(\boldsymbol{Z} - \boldsymbol{X\beta})\right]. \tag{3.7}$$

From (3.7) the conjugacy of the inverse gamma as σ's prior becomes apparent. With an exponential correlation structure, the spatial correlation matrix is $g(\rho, \boldsymbol{C}) = \exp(-\boldsymbol{C}/\rho)$ and so (3.7) can be written as

$$p(\boldsymbol{\theta}|\boldsymbol{\beta}, \boldsymbol{Z}) \propto p(\rho)\sigma^{-\alpha-1-n/2}|e^{-\frac{\boldsymbol{C}}{\rho}}|^{-\frac{1}{2}} \exp\left[-\frac{1}{\sigma}\left(\frac{1}{\gamma} + \frac{1}{2}(\boldsymbol{Z} - X\beta)^{\mathrm{T}}(e^{-\frac{\boldsymbol{C}}{\rho}})^{-1}(\boldsymbol{Z} - X\beta)\right)\right].$$

Thus the full conditional distribution of σ also has an inverse gamma with parameter

$$\begin{aligned}\tilde{\alpha} &= \alpha + n/2 \\ \tilde{\gamma} &= \left(\frac{1}{\gamma} + \frac{1}{2}(\boldsymbol{Z} - X\beta)^{\mathrm{T}}(e^{-\frac{\boldsymbol{C}}{\rho}})^{-1}(\boldsymbol{Z} - X\beta)\right)^{-1}.\end{aligned}$$

However ρ does not have a standard full conditional distribution, forcing us to use Metropolis-Hasting algorithm proposed by Hastings (1970) to sample ρ from (3.7) with σ fixed.

If σ_e^2 and σ_δ^2 have inverse gamma priors with parameters (shape=α_1,scale=$1/\lambda_1$) and (shape=α_2,scale=$1/\lambda_2$) respectively, the full conditional distributions of

σ_e^2 and σ_δ^2 become

$$\begin{aligned}
&\sigma_\delta^2|(a, b, \tilde{\boldsymbol{Z}}, \boldsymbol{Z}) \sim \mathrm{IG}\left(\text{shape} = \alpha_1 + \mathrm{m}/2, \text{scale} = (\lambda_1 + \frac{1}{2}\tilde{\lambda})^{-1}\right) \\
&\text{with } \tilde{\lambda} = (\tilde{\boldsymbol{Z}} - a - bA_1\boldsymbol{Z})^{\mathrm{T}}(\tilde{\boldsymbol{Z}} - a - bA_1\boldsymbol{Z}) \text{ and} \\
&\sigma_e^2|(\hat{\boldsymbol{Z}}, \boldsymbol{Z}_2) \sim \mathrm{IG}\left(\text{shape} = \alpha_2 + \mathrm{n}/2, \text{scale} = (\lambda_2 + \frac{1}{2}\gamma)^{-1}\right) \\
&\text{with } \gamma = (\hat{\boldsymbol{Z}} - A_0\boldsymbol{Z})^{\mathrm{T}}(\hat{\boldsymbol{Z}} - A_0\boldsymbol{Z}).
\end{aligned} \tag{3.8}$$

Letting the prior for a, b be

$$\begin{pmatrix} a \\ b \end{pmatrix} \sim \mathrm{MVN}(\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{F}})$$

yields the full conditional distribution of the calibration parameters a, b as

$$\begin{pmatrix} a \\ b \end{pmatrix} |\tilde{\boldsymbol{Z}}, A_1\boldsymbol{Z}, \boldsymbol{\Sigma}_2 \sim \mathrm{MVN}(\bar{\boldsymbol{B}}\bar{\boldsymbol{C}}, \bar{\boldsymbol{B}}),$$

where

$$\bar{\boldsymbol{B}} = \begin{pmatrix} \mathbf{1}^{\mathrm{T}} \\ (A_1\boldsymbol{Z})^{\mathrm{T}} \end{pmatrix} (\boldsymbol{\Sigma}_2)^{-1} \begin{pmatrix} \mathbf{1}, & A_1\boldsymbol{Z} \end{pmatrix} + \bar{\boldsymbol{F}}^{-1}$$

and

$$\bar{\boldsymbol{C}} = \begin{pmatrix} \mathbf{1}^{\mathrm{T}} \\ (A_1\boldsymbol{Z})^{\mathrm{T}} \end{pmatrix} (\boldsymbol{\Sigma}_2)^{-1} \begin{pmatrix} \mathbf{1}, & \tilde{\boldsymbol{Z}} \end{pmatrix} + \bar{\boldsymbol{F}}^{-1}\bar{\boldsymbol{\beta}}.$$

We define $\mathbf{1}$ to be an column vector having the same dimension as $\tilde{\boldsymbol{Z}}$.

3.2 Bayesian Melding in Non-stationary Case

The previous section presents the full conditional distributions of all the parameters for a stationary true underlying process distribution. The full conditional distributions are the same for the non-stationary case except for the spatial covariance parameters (σ and ρ). In model (2.8), the priors for the "main-effects" (α, r_i, c_j and α', r_i', c_j') are independent multivariate normal distributions with means $\boldsymbol{\mu}_\sigma$ and $\boldsymbol{\mu}_\rho$.

Fuentes and Smith (2001) propose the non-stationary model (2.8) without giving much detail about fitting the model. In this section, we derive the full conditional distribution of $\boldsymbol{\sigma}$, $\boldsymbol{\mu}_\sigma$, τ_σ and η_σ. For that purpose we let $\boldsymbol{Y}$ be the design matrix in model (2.8). The parameters associated with the *sill* include $\boldsymbol{\sigma}$, $\boldsymbol{\mu}_\sigma$, τ_σ and η_σ, while those associated with the *range* include $\boldsymbol{\rho}$, $\boldsymbol{\mu}_\rho$, τ_ρ and η_ρ. Because the full conditional distribution of the parameters associated with the *sill* are analogous to those associated with the *ranges*, we only give the derivation of the full conditional distributions of parameters associated with the former. In that derivation, H represents all the other parameters as well as the available data. The full conditional distribution of $\boldsymbol{\sigma}$ is the following:

$$p(\boldsymbol{\sigma}|H) \propto |\boldsymbol{\Sigma}(\boldsymbol{\sigma},\boldsymbol{\rho})|^{-\frac{1}{2}}|\boldsymbol{\Sigma}(\tau_\sigma,\eta_\sigma)|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\left[(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Sigma}(\boldsymbol{\sigma},\boldsymbol{\rho})^{-1}(\boldsymbol{Z}-\boldsymbol{X}\boldsymbol{\beta})\right]\}$$
$$\exp\{-\frac{1}{2}\left[(\boldsymbol{\sigma}-\boldsymbol{Y}\boldsymbol{\mu}_\sigma)^{\mathrm{T}}\boldsymbol{\Sigma}(\tau_\sigma,\eta_\sigma)^{-1}(\boldsymbol{\sigma}-\boldsymbol{Y}\boldsymbol{\mu}_\sigma)\right]\},$$

where $\boldsymbol{\Sigma}(\boldsymbol{\sigma}, \boldsymbol{\rho})$, the spatial covariance matrix of $\boldsymbol{Z}$ and $\boldsymbol{\Sigma}(\tau_\sigma, \eta_\sigma)$ is the spatial covariance of the vector $\boldsymbol{\sigma}$ for the latent stationary processes. In the non-stationary case, the full conditional distribution of $\boldsymbol{\sigma}$ is no longer a standard distribution, necessitating use of the Metropolis-Hasting algorithm to update $\boldsymbol{\sigma}$ as a block.

Since $\boldsymbol{\mu}_\sigma$ has a normal prior with mean $\boldsymbol{\beta}_\sigma$ and variance matrix $\boldsymbol{F}_\sigma$, its full conditional distribution is

$$\begin{aligned}
&\boldsymbol{\mu}_\sigma|\boldsymbol{\sigma} \sim \mathrm{MVN}(Bb, B) \\
&B^{-1} = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\Sigma}(\tau_\sigma, \eta_\sigma)^{-1}\boldsymbol{Y} + \boldsymbol{F}_\sigma^{-1} \\
&b = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\Sigma}(\tau_\sigma, \eta_\sigma)^{-1}\boldsymbol{\sigma} + F_\sigma^{-1}\boldsymbol{\beta}_\sigma.
\end{aligned}$$

The full conditional distribution of τ_σ is

$$p(\tau_\sigma|\boldsymbol{Z}, \eta_\sigma) \propto p(\tau_\sigma, \eta_\sigma)|\boldsymbol{\Sigma}(\tau_\sigma, \eta_\sigma)|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{\sigma} - \boldsymbol{Y}\boldsymbol{\mu}_\sigma)^{\mathrm{T}}\boldsymbol{\Sigma}(\tau_\sigma, \eta_\sigma)^{-1}(\boldsymbol{\sigma} - \boldsymbol{Y}\boldsymbol{\mu}_\sigma)\right].$$

Our Gibbs sampling algorithm must include Metropolis-Hasting steps to sample from the full conditional distributions whenever they cannot be specified in a closed form.

3.3 MCMC Algorithm

To sample from the posterior distribution of the parameters in the BM model (2.9), we use the Gibbs sampling algorithm proposed by Gelfand and Smith

(1990) as implemented by Fuentes and Raftery (2005). First, we choose some arbitrary initial values for $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ as $\boldsymbol{\theta}^{(1)}$, $\boldsymbol{\beta}^{(1)}$. Then, the Gibbs sampling is implemented in the following three stages.

- **Stage 1:** Given all the other parameters and the available data, realizations of the true underlying process $\{\boldsymbol{Z}(\boldsymbol{s})\}$ are updated at n stations and L sampling points within each of the m grid cells. In this stage, a random sample of $\boldsymbol{Z}$ is generated by using (3.4).

- **Stage 2:** First $\boldsymbol{\theta} = (\sigma, \rho)$ is updated given $\boldsymbol{\beta}$ and $\boldsymbol{Z}$ obtained in Stage 1. Second, $\boldsymbol{\beta}$ given $\boldsymbol{\theta}$ and $\boldsymbol{Z}$ is updated by (3.5). Updating $\boldsymbol{\beta}$ and σ is easy because their full conditional distributions are multivariate normal and inverse gamma respectively. However, the full conditional distribution of ρ does not have a closed form, so we have to use Metropolis-Hasting algorithm to update it.

- **Stage 3:** In this stage, $\sigma_e^2, \sigma_\delta^2$ and a, b are easily updated given all the other parameters since their full conditional distributions are either normal or inverse gamma. So, it is easy to update them.

The above Gibbs sampling algorithm is nearly identical for the stationary and non-stationary cases, the only difference being in the updating of $\boldsymbol{\theta}$ in Stage 2.

3.4 Spatial Prediction

This section indicates how the BM model can be used to predict realizations of the underlying process at unmonitored sites using the available data.

Denote by $\boldsymbol{Z}_u$, realizations of the true underlying process at the unmonitored sites of interest. Finding the interpolation procedure entails finding the posterior distribution of $\boldsymbol{Z}_u|\boldsymbol{Z}_g$, $\boldsymbol{Z}_g$ being realizations of the true underlying process at monitoring stations and sampling points within grid cells. Observe that $\mathrm{E}(\boldsymbol{Z}_g) = \boldsymbol{\mu}_g$ and $\mathrm{E}(\boldsymbol{Z}_u) = \boldsymbol{\mu}_u$. Thus the conditional mean and variance of $\boldsymbol{Z}_u|\boldsymbol{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}$ are

$$\begin{aligned}
\mathrm{E}(\boldsymbol{Z}_u|\boldsymbol{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}) &= \boldsymbol{\mu}_u + \boldsymbol{\Sigma}_{ug}\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{Z}_g - \boldsymbol{\mu}_g), \text{ and} \\
\mathrm{Var}(\boldsymbol{Z}_u|\boldsymbol{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}) &= \boldsymbol{\Sigma}_u - \boldsymbol{\Sigma}_{ug}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Sigma}_{gu}.
\end{aligned}$$

The covariance matrix of $\boldsymbol{Z}_u$ is $\boldsymbol{\Sigma}_u$ and the covariance matrix between $\boldsymbol{Z}_u$ and $\boldsymbol{Z}_g$ is $\boldsymbol{\Sigma}_{ug}$. Let $\boldsymbol{\Sigma}_{gu}$ denote the transpose of $\boldsymbol{\Sigma}_{ug}$. In the BM model (2.9), we have $\hat{Z}(\boldsymbol{s}) = Z(\boldsymbol{s}) + e(\boldsymbol{s})$. So, the conditional mean and variance of $\hat{\boldsymbol{Z}}_u|\boldsymbol{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}$ are

$$\begin{aligned}
\mathrm{E}(\hat{\boldsymbol{Z}}_u|\boldsymbol{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}) &= \boldsymbol{\mu}_u + \boldsymbol{\Sigma}_{ug}\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{Z}_g - \boldsymbol{\mu}_g), \text{ and} \\
\mathrm{Var}(\hat{\boldsymbol{Z}}_u|\boldsymbol{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}) &= \boldsymbol{\Sigma}_u - \boldsymbol{\Sigma}_{ug}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Sigma}_{gu} + \sigma_e^2\boldsymbol{I},
\end{aligned}$$

in which the dimension of the identity matrix $\boldsymbol{I}$ is the number of stations to be predicted.

The marginal distribution of $\hat{\boldsymbol{Z}}_u|\hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}}$ is

$$p(\hat{\boldsymbol{Z}}_u|\hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}}) = \int p(\boldsymbol{Z}_u|\boldsymbol{\phi})p(\boldsymbol{\phi}|\hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}})d\boldsymbol{\phi}, \tag{3.9}$$

in which $\boldsymbol{\phi}$ stands for all the other parameters such as $\boldsymbol{Z}_g$, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ that can be approximated by

$$p(\hat{\boldsymbol{Z}}_u|\boldsymbol{Z}_g) \approx \frac{1}{n}\sum_{i=1}^{n} p(\hat{\boldsymbol{Z}}_u|\boldsymbol{Z}_g, \boldsymbol{\phi}^i),$$

where $\boldsymbol{\phi}^i (i = 1, \cdots, n)$ is the i-th MCMC sample of all the parameters.

3.5 Extension to Ensembles

Previous sections consider the problem of combining measurements from monitoring sites made with a single instrument with output from just one deterministic model. In fact, the BM model can be extended to combine data from ensembles of measuring instruments and of deterministic models. For simplicity we consider the case of just a single measuring instrument and multiple deterministic models although the extension to incorporate multiple measuring instruments will be obvious.

To extend model (2.9), suppose $\tilde{\boldsymbol{Z}}_1, \cdots, \tilde{\boldsymbol{Z}}_p$ are output from an ensemble of p deterministic models. To ensure a non-singular spatial covariance matrix, suppose no overlap between the monitoring sites and sampling points within each grid cell for all the different deterministic models. For simplicity, we

assume the number of sampling points within each grid cell is the same, L, for all the deterministic models although extension to differing numbers is straightforward. We also assume each deterministic model has model outputs on m grid cells. Then the dimension of $\boldsymbol{Z}$ is $n + p \times m \times L$. As in (3.1), the joint posterior distribution can be decomposed as follows:

$$
\begin{aligned}
& p(\hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}}_1, \cdots, \tilde{\boldsymbol{Z}}_p, \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{\theta}, a_1, \cdots, a_p, b_1, \cdots, b_p, \sigma_e^2, \sigma_{\delta,1}^2, \cdots, \sigma_{\delta,p}^2) \\
= \; & p(\hat{\boldsymbol{Z}} | \boldsymbol{Z}, \sigma_e^2) \prod_{i=1}^{p} p(\tilde{\boldsymbol{Z}}_i | \boldsymbol{Z}, a_i, b_i, \sigma_{\delta,i}^2) p(\boldsymbol{Z} | \boldsymbol{\beta}, \boldsymbol{\theta}) p(\sigma_e^2, \sigma_{\delta,1}^2, \cdots, \sigma_{\delta,p}^2, \boldsymbol{\beta}, \boldsymbol{\theta}) \\
= \; & \Phi_{\boldsymbol{\Sigma}_0}(\hat{\boldsymbol{Z}} - \boldsymbol{Z}_0) \prod_{i=1}^{p} \Phi_{\boldsymbol{\Sigma}_i}(\tilde{\boldsymbol{Z}}_i - a_i - b_i A_i \boldsymbol{Z}) p(\sigma_{\delta,i}^2) \Phi_{\boldsymbol{\Sigma}(\boldsymbol{\theta})}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta}) p(\sigma_e^2) p(\boldsymbol{\theta}) p(\boldsymbol{\beta}).
\end{aligned}
$$

Deterministic model i has calibration parameters a_i and b_i, model output error variance parameter $\sigma_{\delta,i}^2$. The covariance matrix of measurement error $\boldsymbol{e}$ is $\boldsymbol{\Sigma}_0 = \sigma_e^2 \boldsymbol{I}$ and the covariance matrix of the model output error vector $\boldsymbol{\delta}_i$ of the i-th deterministic model is $\boldsymbol{\Sigma}_i = \sigma_{\delta,i}^2 \boldsymbol{I}$. The dimension of matrix A_0 is $n \times (n + p \times m \times L)$ and that of matrix A_i $(i = 1, \cdots, p)$ is $m \times (n + p \times m \times L)$. The first n columns of matrix A_0 form an identity matrix and the remaining elements are all 0. Row j of matrix A_i $(i = 1, \cdots, p)$ has elements $\frac{1}{L}$ from column $n + 1 + (i-1) \times m \times L + (j-1) \times L$ to $n + 1 + (i-1) \times m \times L + j \times L$ and all other elements are zero.

The density of the conditional distribution of $(\hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}}_1, \cdots, \tilde{\boldsymbol{Z}}_p, \boldsymbol{Z})$ given

$(\boldsymbol{\beta}, \boldsymbol{\theta}, a_1, \cdots, a_p, b_1, \cdots, b_p, \sigma_e^2, \sigma_{\delta,1}^2, \cdots, \sigma_{\delta,p}^2)$ is

$$
\begin{aligned}
& p(\hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}}_1, \cdots, \tilde{\boldsymbol{Z}}_p, \boldsymbol{Z}) | \boldsymbol{\beta}, \boldsymbol{\theta}, a_1, \cdots, a_p, b_1, \cdots, b_p, \sigma_e^2, \sigma_{\delta,1}^2, \cdots, \sigma_{\delta,p}^2) \\
\propto \; & \exp\left\{-\frac{1}{2}\left[\left(\hat{\boldsymbol{Z}} - A_0\boldsymbol{Z}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_1^{-1}\left(\hat{\boldsymbol{Z}} - A_0\boldsymbol{Z}\right)\right.\right. \\
+ \; & \sum_{i=1}^{p}\left(\tilde{\boldsymbol{Z}}_i - a_i - b_i A_i \boldsymbol{Z}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_i^{-1}\left(\tilde{\boldsymbol{Z}}_i - a_i - b_i A_i \boldsymbol{Z}\right) \\
+ \; & \left.\left.(\boldsymbol{Z} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu})\right]\right\} \\
= \; & \exp\left\{-\frac{1}{2}\left[-\boldsymbol{Z}^{\mathrm{T}}\left(A_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\hat{\boldsymbol{Z}} + \sum_{i=1}^{p} b_i A_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\tilde{\boldsymbol{Z}}_i - a_i) + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)\right.\right. \\
- \; & \left(\hat{\boldsymbol{Z}}^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}A_0 + \sum_{i=1}^{p} b_i(\tilde{\boldsymbol{Z}}_i - a_i)\boldsymbol{\Sigma}_i^{-1}A_i + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{Z} \\
+ \; & \left.\left.\boldsymbol{Z}^{\mathrm{T}}\left(A_1^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}A_1 + \sum_{i=1}^{p} b_i^2 A_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}A_i + \boldsymbol{\Sigma}^{-1}\right)\boldsymbol{Z}\right]\right\} + C, \qquad (3.10)
\end{aligned}
$$

$\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ being the mean and variance of $\boldsymbol{Z}$, C, the other terms without $\boldsymbol{Z}$.

Like (3.4), the full conditional distribution of $\boldsymbol{Z}$ is the following.

$$
\begin{aligned}
& \boldsymbol{Z} | \text{ other parameters and all data} \sim \mathrm{MVN}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}) \\
& \tilde{\boldsymbol{\Sigma}}^{-1} = (A_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}A_0 + \sum_{i=1}^{p} b_i^2 A_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}A_i + \boldsymbol{\Sigma}^{-1}) \\
& \tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\Sigma}}\left(A_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\hat{\boldsymbol{Z}} + \sum_{i=1}^{p} b_i A_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\tilde{\boldsymbol{Z}}_i - a_i) + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right).
\end{aligned}
$$

The prior distribution of (a_i, b_i), $i = 1, \cdots, p$, is

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim \mathrm{MVN}(\bar{\boldsymbol{\beta}}_i, \bar{\boldsymbol{F}}).$$

Their full conditional distribution is

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} | \tilde{\boldsymbol{Z}}_i, \boldsymbol{Z}, \boldsymbol{\Sigma}_i \sim \mathrm{MVN}(\bar{\boldsymbol{B}}\bar{\boldsymbol{C}}, \bar{\boldsymbol{B}}),$$

where

$$\bar{\boldsymbol{B}} = \begin{pmatrix} \mathbf{1}^{\mathrm{T}} \\ (A_i \boldsymbol{Z})^{\mathrm{T}} \end{pmatrix} (\boldsymbol{\Sigma}_i)^{-1} \begin{pmatrix} \mathbf{1}, & A_i \boldsymbol{Z} \end{pmatrix} + \bar{\boldsymbol{F}}_i^{-1}$$

and

$$\bar{\boldsymbol{C}} = \begin{pmatrix} \mathbf{1}^{\mathrm{T}} \\ (A_i \boldsymbol{Z})^{\mathrm{T}} \end{pmatrix} (\boldsymbol{\Sigma}_i)^{-1} \begin{pmatrix} \mathbf{1}, & \tilde{\boldsymbol{Z}}_i \end{pmatrix} + \bar{\boldsymbol{F}}_i^{-1} \bar{\boldsymbol{\beta}}_i.$$

3.6 Reversible Jump MCMC

Previous sections assume a known degree in the polynomial representing the mean function $\mu(\boldsymbol{s}) = \boldsymbol{X}(\boldsymbol{s})\boldsymbol{\beta}$. In other words, the coefficient vector $\boldsymbol{\beta}$ has known dimension. Green (1995) proposes the reversible jump MCMC to allow the dimension of $\boldsymbol{\beta}$ to be an unknown parameter. For the BM model, we let k be $\boldsymbol{\beta}$'s dimension and $\boldsymbol{\eta}_k = (\boldsymbol{\beta}_k, \boldsymbol{\theta})$. The reversible jump MCMC allows k to vary, while the dimension of all other parameters are fixed. The

objective: to sample from the joint posterior distribution of k and $\boldsymbol{\eta}_k$, that is, $p(k, \boldsymbol{\eta}_k | data)$. To achieve that, we first find the full joint conditional distribution of k and $\boldsymbol{\eta}_k$. Given $\boldsymbol{Z}$, that full conditional distribution is

$$p(k, \boldsymbol{\eta}_k | \boldsymbol{Z}) \propto p(\boldsymbol{Z}, \boldsymbol{\eta}_k, k) = p(\boldsymbol{Z} | k, \boldsymbol{\eta}_k) p(\eta_k | k) p(k),$$

$p(k)$ being the prior distribution of the dimension k and $p(\eta_k | k)$, the prior distribution of $\boldsymbol{\eta}_k$ given k. Suppose initially the dimension of $\boldsymbol{\beta}$ is p ($p > 0$). Then the next iteration has two possible outcomes for k: increase the dimension by one (a "birth") or decrease it by one (a "death"). The probability of a "birth" or "death" is 1/2. The jump scheme can be quite arbitrarily but we make the above choice because we favor "birth" or "death" equally. After choosing either "birth" or "death", we accept/reject the jump with probability $\alpha/(1-\alpha)$, the specification of α, being explained below. If we reject the jump, then we stay with the current dimension p; otherwise, the dimension k will be either $p+1$ or $p-1$. In the same iteration, the next step updates the parameters $\boldsymbol{\eta}$ given k as in the fixed dimensional case. By using the formulas provided in Green (1995), the acceptance probabilities of "birth" and "death" are the following.

- "Birth". To jump from the previous dimension $k = p$ to a new dimension $k^* = p + 1$, we need to propose one extra coefficient β_{new} for one extra covariate. That extra coefficient β_{new} is proposed by a proposal distribution with density $q(\cdot)$. Because β_{new} is proposed independently

of other coefficients the Jacobian of transforming from $\boldsymbol{\beta}$ to $\boldsymbol{\beta}^*$ is 1. So, the new coefficient vector is $\boldsymbol{\beta}^* = (\boldsymbol{\beta}, \beta_{new})$ and the acceptance probability of $k^* = p + 1$ is

$$\alpha = min\{1, \frac{p(k^*)p(\boldsymbol{\beta}^*|k^*)p(\boldsymbol{Z}|k^*, \boldsymbol{\beta}^*)}{p(k)p(\boldsymbol{\beta}|k)p(\boldsymbol{Z}|k, \boldsymbol{\beta})q(\beta_{new})}\}.$$

- "Death". To jump from the previous dimension $k = p$ to a new dimension $k^* = p - 1$, we need to delete one coefficient from the coefficient vector. We choose to delete the last coefficient of the vector $\boldsymbol{\beta}$. The reason is that we arrange the covariates from lower order of the coordinates to higher order and usually the higher order of the coordinates are more likely to be insignificant than lower ones. Again, the Jacobin is 1 and the acceptance probability of $k^* = p - 1$ is

$$\alpha = min\{1, \frac{p(k^*)p(\boldsymbol{\beta}^*|k^*)p(\boldsymbol{Z}|k^*, \boldsymbol{\beta}^*)q(\beta_p)}{p(k)p(\boldsymbol{\beta}|k)p(\boldsymbol{Z}|k, \boldsymbol{\beta})}\}.$$

After choosing dimension k, we update all other parameters by using Gibbs sampling as in the fixed dimension case. In Section 4.3, we conduct a simulation to see how well the reversible jump MCMC works when incorporated into the BM model.

The discussion above shows how to use the reversible jump MCMC algorithm to choose the dimension of $\boldsymbol{\beta}$. That algorithm can also be used to choose the dimension of coefficients of the additive calibration $a(\boldsymbol{s})$. If we

let $a(\boldsymbol{s}) = \boldsymbol{Y}(\boldsymbol{s})\boldsymbol{\beta}_a$, $\boldsymbol{Y}(\boldsymbol{s})$ being the polynomial function of the coordinate at location $\boldsymbol{s}$, then we can use reversible jump MCMC to choose the dimension of coefficients $\boldsymbol{\beta}_a$. The detailed algorithm is very similar to the above when reversible jump MCMC is applied to choose the dimension of $\boldsymbol{\beta}$. The current software does not include the reversible jump MCMC to choose the dimension of coefficients $\boldsymbol{\beta}_a$, but we are planning to incorporate that into the software in the near future.

3.7 Bayesian Melding Model Software

We wrote a BM model program in R, developed by the R Development Core Team (2006). The R program is online at http://enviro.stat.ubc.ca/melding/meldingcode.zip. The code for the stationary BM model is in the directory "mcmc" and the code for stationary BM model incorporating the reversible jump MCMC is in the directory "rjmcmc". The code for the non-stationary BM model will be available online soon.

The subdirectory "debug" in directory "mcmc" is used to debug the R programs. That directory includes the file "simudata.s" simulating the measurements and model output, "mcmc.s", the main function to implement the MCMC algorithm, "update.s", various functions to sample parameters from their full conditional distributions. The description of the main function "melding" in file "mcmc.s" is the following.

Dependence: R≥2.3.0 and packages "MASS".

Usage: melding(m1,m2,nm,sam.sloc,sam.sloc1,ton,burnin,zhat,Zbtilde,degree,cov.model)

Arguments:

- nm: number of sampling points in each grid cell of the model output.
- sam.sloc: coordinates of the sampling points and the monitoring stations. It should be a $(m2 \times nm + m1) \times 2$ matrix. The first $m2 \times nm$ rows are the coordinates of the sampling points in grid cells. The last $m1$ rows are the coordinates of the monitoring stations.
- sam.sloc1: coordinates of the unmonitored stations where the measurements are to be predicted.
- ton: number of MCMC iterations in the Gibbs sampling.
- burnin: the "burn-in" period of the Gibbs sampling.
- zhat: the measurements vector.
- Zbtilde: the model output vector.
- degree: degree of the polynomial function $f(\cdot)$ for the mean $\mu(\boldsymbol{s}) = f(\boldsymbol{s})\boldsymbol{\beta}$ and the following options are allowed:

  0 the mean is assumed constant across space.

  1 the mean is assumed to be a first order polynomial on the coordinates: $\mu(\boldsymbol{s}) = \beta_0 + \beta_1 s_1 + \beta_2 s_2$.

2 the mean is assumed to be a second order polynomial on the coordinates: $\mu(\boldsymbol{s}) = \beta_0 + \beta_1 s_1 + \beta_2 s_2 + \beta_3 s_1^2 + \beta_4 s_2^2 + beta_5 s_1 s_2$.

- cov.model: a string with the name of the correlation function. The options are one of the following three functions:

  "exponential": $\exp(-d/\rho)$, d is the distance and rho is the range parameter.

  "Gaussian": $\exp(-d^2/\rho)$, d is the distance and rho is the range parameter.

  "Matern": $[(2^\nu - 1)\Gamma(\nu)]^{-1} (d/\rho)^\nu K_\nu(d/\rho)$. ν is the smoothing parameter and $K_\nu(\cdot)$ denotes the modified Bessel function of the third kind of order ν.

The function "melding" will return a list of the following objects:

- "beta.est": posterior mean of the coefficient vector $\boldsymbol{\beta}$.
- "beta.est.sd": posterior standard deviation of the coefficient vector $\boldsymbol{\beta}$.
- "theta.est": posterior mean of the spatial correlation vector $\boldsymbol{\theta}$.
- "theta.est.sd": posterior standard deviation of the spatial correlation vector $\boldsymbol{\theta}$.
- "prediction": posterior mean of the spatial prediction.
- "pred.q1": 5% quantile of the posterior distribution of the spatial prediction.

- "pred.q2": 95% quantile of the posterior distribution of the spatial prediction.

- "ab.est": posterior mean of the additive and multiplicative parameters a and b.

- "ab.est.sd": posterior standard deviation of the additive and multiplicative parameters a and b.

- "sigmae.est": posterior mean of the measurement error variance parameter σ_e^2.

- "sigmae.est.sd": posterior standard deviation of the measurement error variance parameter σ_e^2.

- "sigmad.est": posterior mean of the model output error variance parameter σ_δ^2.

- "sigmad.est.sd": posterior standard deviation of the model output error variance parameter σ_δ^2.

The description of various functions in file "update.s" is the following.

- "updatebeta(...)": generate MCMC sample from the full conditional distribution of $\boldsymbol{\beta}$. The arguments of this function include

  y: the realizations of the true underlying process $\{\boldsymbol{Z}(\boldsymbol{s})\}$ at monitored stations and sampling points within grid cells.

x: the covariate matrix at monitored stations and sampling points within grid cells.

prior.mean: prior mean of the coefficient vector $\boldsymbol{\beta}$.

prior.var.solve: inverse of the prior variance matrix of the coefficient vector $\boldsymbol{\beta}$.

sigma.solve: inverse of the spatial covariance matrix of true underlying process $\{\boldsymbol{Z}\}$.

- "updatetheta(...)": generate MCMC sample from the full conditional distribution of $\boldsymbol{\theta}$.The arguments of this function include

  diff: the residuals of the true underlying process.

  theta: the values of the covariance parameters $\boldsymbol{theta}$ from the previous MCMC iteration.

  Distance: the Euclidean distance matrix between all the locations including monitored stations and sampling points within grid cells.

  n: number of monitored stations and sampling points within grid cells.

  cov.model: one of the three possible choices for the covariance function: "exponential", "Gaussian" and "matern".

- "updatesigma(...)": generate MCMC sample from the full conditional distribution of σ_δ^2 and σ_e^2. The arguments of this function include

  zhat: the measurements vector.

Zbtilde: the model output vector.

nm: number of sampling points in each grid cell of the model output; n: number of monitored stations and sampling points within grid cells; m2: number of grid cells.

a: additive bias parameter a; b: multiplicative bias parameter b.

A2: matrix A_1 as described in Section 3.1.

- "updateab(...)":generate MCMC sample from the full conditional distribution of a and b. The arguments of this function include

  Zbtilde: same as in the function "updatesigma(...)". ,y,ab0,fb.solve,nm,m2,A2.

  sigmad: the output error variance parameter σ_δ^2.

  y: same as in the function "updatesigma(...)".

  n,m2,A2: same as in the function updatesigma(...).

- "updatezs(...)": generate MCMC sample from the full conditional distribution of $\boldsymbol{Z}$. The arguments of this function include

  X: the covariate matrix at monitored stations and sampling points within grid cells.

  zhat, Zbtilde, nm,m2: same as in the function "updatesigma(...)".

  sigmad, sigmae: the output error variance parameter σ_δ^2 and measurement error variance parameter σ_e^2.

sigma3.solve: inverse of the spatial covariance matrix of true underlying process $\{\boldsymbol{Z}\}$.

Reversible Jump MCMC is used to do the variable selection in a Bayesian framework. It is functioning similar to the step-wise regression in classical framework. All the functions in "update.s" are the same as in the BM model described above. In the function "rjmelding(...)", we do not specify the "degree", which is estimated by the reversible jump MCMC. The function "rjmeding(...)" only returns the dimension of the coefficient $\boldsymbol{\beta}$ in the mean of underlying true process. More details of the software can be found in Liu (2007).

Chapter 4

Testing the Bayesian Melding Model

This chapter presents an extensive simulation study of the BM model under a variety of scenarios that assess the method while checking the R code developed. We begin with the stationary BM model, one deterministic model and one measuring instrument. Then we move on to the case with an ensemble of deterministic models. We include a simulation study that incorporates a reversible jump MCMC into the BM model. In addition, studies are devoted to several other issues such as how the estimation of covariance parameters can be improved by better monitored site layouts and the effect of choosing bandwidth h in non-stationary BM model.

4.1 A Single Deterministic Model

This simulation, which assumes a stationary true underlying process, has two purposes. First, it validates the MCMC algorithm for implementing the BM model. Second, it compares the predictions of Kriging with those of the BM

model.

Simulation Settings

In the simulation, we have 20 monitored sites and 100 un-monitored sites whose responses are to be predicted. The number of available grid cells increases from 0 to 50 through the sequence: 0, 2, 10, 20, 30, 50. To reduce our computational burden, we only have one sampling point within each grid cell. The coordinates of the available/unavailable stations and sampling points within grid cells are generated uniformly on $[-5, 5] \times [-5, 5]$. Figure 4.1 depicts the locations of sites (monitored and un-monitored) and sampling points. For each combination of sites and sampling points, we generate 50 independent dataset of the measurements and model outputs according to the BM model (2.9).

The measurements generated at the 100 unmeasured stations will not be used in the MCMC prediction but left for validation. As the number of grid cells increases, we investigate changes in prediction accuracy and coverage probability for different prediction intervals at various credibility levels (95%, 90%, 80%, 70%, 60%, 40%).

In BM 2.9, we let the mean of the true underlying process be

$$\mathrm{E}\left[Z(\boldsymbol{s})\right] = \mu(\boldsymbol{s}) = \beta_0 + \beta_1 s_1 + \beta_2 s_2, \tag{4.1}$$

s_1, s_2 being the coordinates of location $\boldsymbol{s}$. We assume $\{\boldsymbol{Z}(\boldsymbol{s})\}$ has an expo-

nential covariance structure. At the same time, we give $\boldsymbol{\beta}$ a prior normal distribution with mean $\boldsymbol{\beta}_0 = (1, 1, 1)$ and diagonal covariance matrix with diagonal elements $(100, 100, 100)$. The covariance parameters σ and ρ have independent prior distributions; σ has an inverse gamma prior distribution with shape = 2, scale = 0.2, while ρ has a gamma prior distribution with shape = 3.5, scale = 2. The calibration parameters a and b have independent normal prior distributions with means $\mu_a = 0$, $\mu_b = 1$ and variances $\sigma_a^2 = 100$, $\sigma_b^2 = 100$. The random variables σ_δ^2 and σ_e^2 have identical and independent inverse gamma distributions with shape = 3 and scale = 0.1.

Using the MCMC algorithm to sample from the joint posterior distribution of all parameters, we take 1000 Gibbs sampling iterations (the first 100 being for "burn-in"), which is enough to ensure convergence as shown in Figure 4.3 and Figure 4.4. That is fortunate since computational times are long and necessity for a longer series of iterations would have greatly limited the scope of our simulation study. The software used in our study is available online (http://enviro.stat.ubc.ca/melding/meldingcode.zip) to enable independent verification of our findings.

Simulation Results

Tables 4.1, 4.2 and 4.3 present estimation results for the parameters. Table 4.4 presents the SSPE (sum of squared prediction errors) for the 100 unmeasured stations. Our results lead to the following over observations.

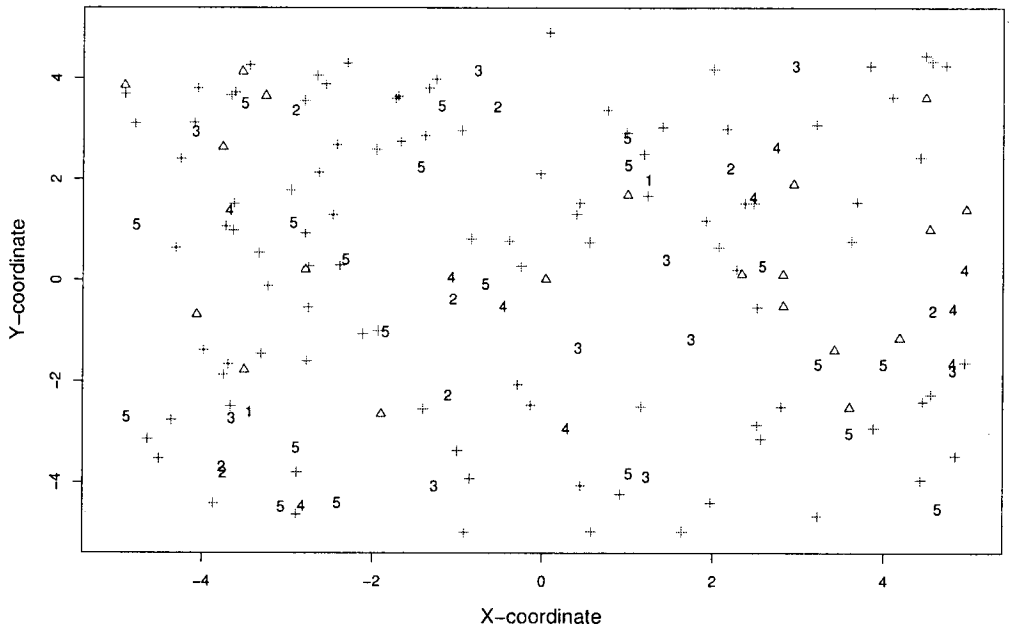

Figure 4.1: Locations of 120 (20 monitored, 100 to be predicted) sites and points sampled from up to 50 grid cells. Monitored sites (Δ), sites with responses to be predicted: $+$, 1: the first 2 points from grid cells, 2: 3-10 points, 3: 11-20 points, 4: 21-30 points, 5: 31-50 points.

- The estimation of a, b and $\boldsymbol{\beta}$ are very close to their true values.

- Based on Figure 4.2, which shows the true underlying process values versus the BM prediction in the case of 50 grid cells, we can see the BM predictions are quite close to the true values.

- The estimates of the covariance parameters $\boldsymbol{\theta} = (\sigma, \rho)$ are reasonably good. However, the *range* parameter ρ proves much more difficult to estimate than the *sill* parameter σ. The posterior distribution in Figure

4.5 shows ρ to be widely dispersed.

- Table 4.4 reveals a significant variation in the SSPE from one simulated dataset to another.

- Table 4.4 shows that adding more grid cells decreases the average SSPE for the BM model. The average is computed over 50 independent dataset. In each, the SSPE does not necessarily decrease in a monotonic fashion as the number of cells increases. However, variation from the non-decreasing pattern in the SSPE may just be due to the sampling variation between the generated dataset. Table 4.5 shows the SSPE obtained by using the true values of the parameters in the predictor. In spite of that advantage, the SSPE for even this predictor does not always decrease monotonically as the number of grid cells increases. (See dataset 4, and 6, for example.)

- Table 4.6 shows that the empirical coverage probability for the BM predictor comes close to the nominal level when we have a reasonably large number of cells (at least 10). However that is not the case when we have just 0 or 2 cells, not surprising given the paucity of data in those situations. Kriging's coverage probability turns out to be much smaller than that for the BM predictor even when no simulated data are available from the deterministic model (so that the two methods compete "head-on"). That result would be anticipated since, as is well known, Kriging underestimates the uncertainty in its spatial predictions.

- Figures 4.3, 4.4 and 4.5 show the MCMC samples from posterior distributions of a, b, β, σ and ρ in the case of 50 grid cells. From these three figures, we can see the Markov chain converges after a few iterations.

Table 4.1: Estimates of a and b with different numbers of model outputs available. The true values are $a = 5.0$ and $b = 2.5$. Columns 2-5 are the averages of the posterior means and standard deviations. The averages are computed over the 50 dataset generated in the simulation study. Since this table is for the estimates of parameters a and b, there is no unit associated to the numbers in this table. Each of these parameters is a scalar.

Number of cells	$\bar{\hat{a}}$	$\bar{sd}$	$\bar{\hat{b}}$	$\bar{sd}$
2	5.61	4.14	2.49	0.22
10	5.11	1.40	2.50	0.09
20	5.24	1.02	2.50	0.08
30	5.12	0.82	2.50	0.07
50	5.15	0.63	2.50	0.06

Improving Covariance Parameter Estimates

Table 4.3 shows that although the posterior mean of ρ gets closer to the true value as the number of grid cells increases from 0 to 50, the posterior standard deviation remains relatively large even with 50 grid cells. That may well be due to insufficient sampling points in close proximity to one another. In other words, we do not have enough information about the small scale process variability needed to accurately estimate the variogram. To explore

Table 4.2: Estimates of the coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ with different numbers of model outputs available. The true values are $\beta_0 = 2.50$, $\beta_1 = 2.90$ and $\beta_2 = 3.20$. Columns 2-7 are the averages of posterior means and standard deviations. The averages are computed over the 50 dataset generated in the simulation study. Since this table is for the estimates of parameters $\beta_0, \cdots, \beta_2$, there is no unit associated to the numbers in this table. Each of these parameters is a scalar.

Number of cells	$\bar{\hat{\beta}}_0$	$\bar{sd}$	$\bar{\hat{\beta}}_1$	$\bar{sd}$	$\bar{\hat{\beta}}_2$	$\bar{sd}$
0	2.35	0.46	2.89	0.12	3.12	0.12
2	2.13	0.75	2.88	0.13	3.11	0.14
10	2.11	0.77	2.88	0.13	3.11	0.14
20	2.09	0.80	2.89	0.13	3.12	0.14
30	2.03	0.82	2.88	0.13	3.12	0.14
50	2.03	0.83	2.88	0.13	3.12	0.14

that conjecture, we carry out a small simulation study with 20 monitoring sites and 50 sampling points in grid cells as in the previous simulation. However, in contrast to the previous case, we concentrate 25 sampling points in a very small region given by $[-0.05, 0.05] \times [-0.05, 0.05]$. Except for that variation, we generate these data in precisely the same way as in the previous simulation. Table 4.7 shows the results for the estimators of the covariance parameters. From that table, we can see that the standard error of estimator for ρ is reduced significantly compared with the result in Table 4.3, thereby adding support to our conjecture.

Table 4.3: Estimates of the covariance parameters in $\boldsymbol{\theta} = (\sigma, \rho)$ with different numbers of model outputs available. The true values are $\sigma = 1.50$, $\rho = 5.00$. Columns 2-5 are the averages of posterior means and standard deviations. The averages are computed over the 50 dataset generated in the simulation study. Since this table is for the estimates of parameters σ and ρ, there is no unit associated to the numbers in this table. Each of these parameters is a scalar.

Number of cells	$\bar{\hat{\sigma}}$	$\bar{sd}$	$\bar{\hat{\rho}}$	$\bar{sd}$
0	1.42	0.26	1.36	1.46
2	1.54	0.31	2.41	2.22
10	1.52	0.35	2.98	2.01
20	1.53	0.33	3.61	2.03
30	1.54	0.33	4.41	2.17
50	1.52	0.27	5.34	2.32

Conclusions

Our simulation results point to strengths and weaknesses in the BM model.

Strengths:

- The BM model can estimate the calibration parameters of the model output very well, given a reasonable number of monitoring sites and grid cells.
- In general, increasing the number of grid cells improves spatial prediction accuracy at un-monitored sites.
- Estimates of the coefficients of the process mean are very good.

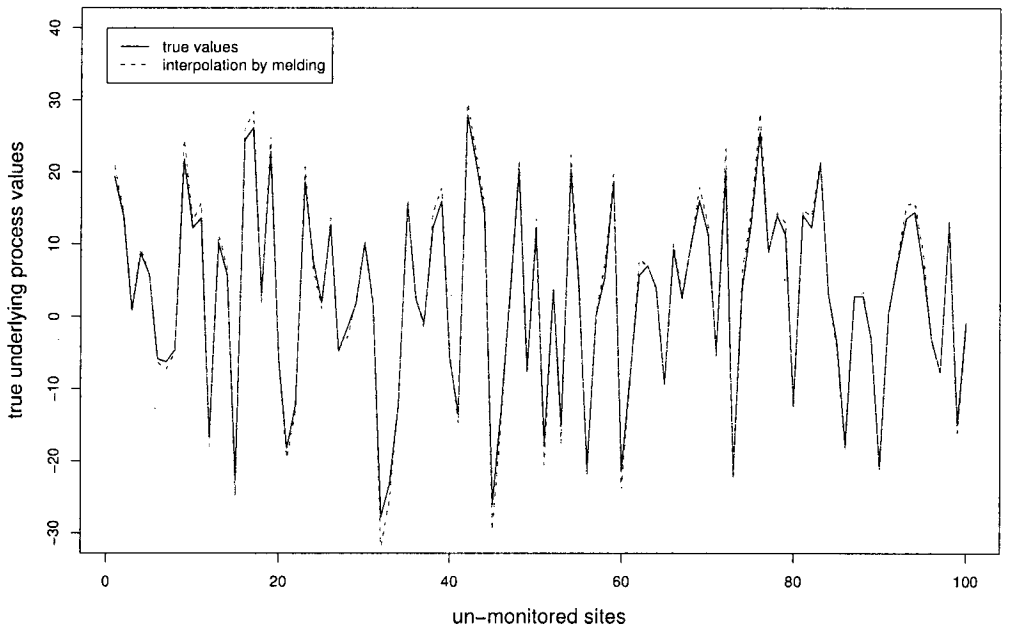

Figure 4.2: Spatial predictions for 100 un-monitored sites by using BM model in the case of 50 grid cells.

- The BM predictor gives much more realistic estimates of the prediction uncertainty than does the classical Kriging approach.

Weaknesses:

- The BM model does not estimate the covariance parameters (σ and ρ) well in terms of their standard errors unless an appreciable number of sampling points are very close to one another. (As seen seen from the previous section, forcing a number of such points into close proximity does decrease their standard errors substantially).

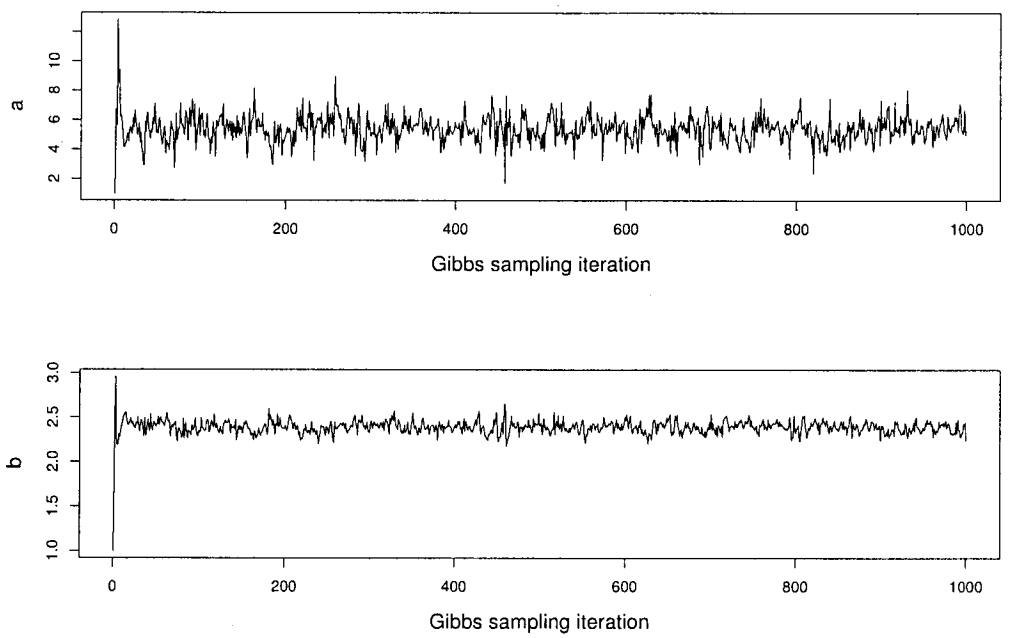

Figure 4.3: MCMC Gibbs samplers of the additive and multiplicative calibration parameters, a and b, respectively, in the case of 50 grid cells.

- The current BM model (2.9) does not include temporal information. For the ozone pollution data studied by Kasibhatla and Chameides (2000), both the real measurements and model output are hourly data, which are temporally correlated. Extending the BM model to embrace random space - time fields would be desirable, as it would enable potentially great strength to be borrowed across time as well as space. One of such extensions is the dynamic BEM model we present later in Chapter 10.

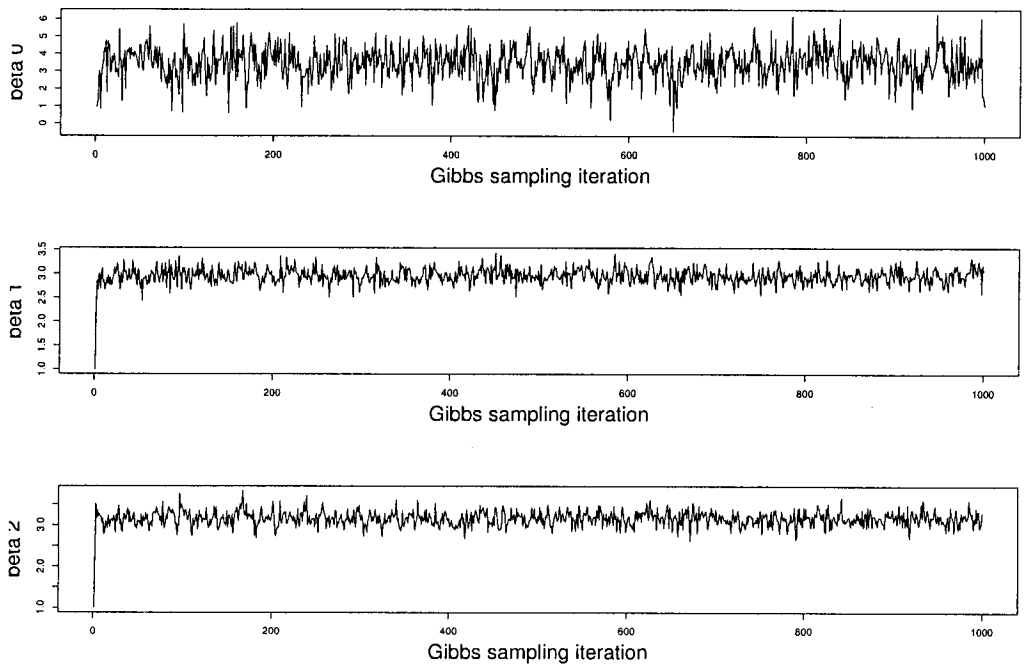

Figure 4.4: MCMC Gibbs samplers of coefficients $\boldsymbol{\beta}$ in the case of 50 grid cells.

- The computational burden imposed by the BM model limits its practicality. That burden stems from the need to invert a large dimensional spatial covariance matrices three times in each Gibbs sampling iteration. For example, 20 monitoring sites and 100 grid cells with 4 sampling points in each leads to a spatial covariance matrix with $20 + 100 \times 4 = 420$ rows and columns. MCMC does not provide a "free Bayesian lunch" as might naively be suggested by its very elegant theory.

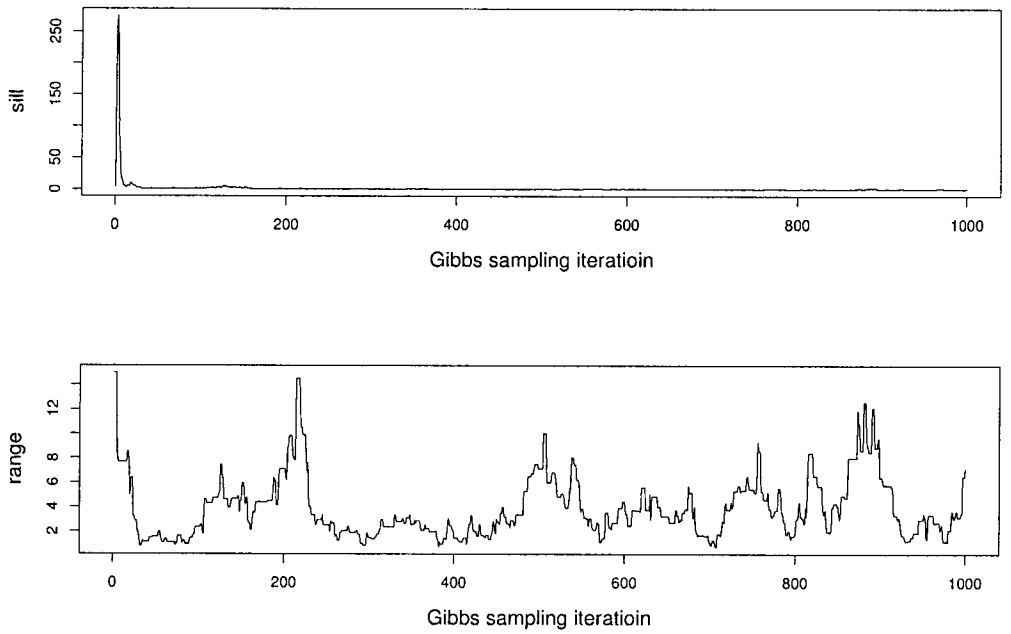

Figure 4.5: MCMC Gibbs samplers of covariance parameters $\boldsymbol{\theta} = (\sigma, \rho)$ in the case of 50 grid cells.

4.2 Ensemble Studies

Simulation settings

This section presents a simulation study illustrating the use of BEM model to combine measurements with model outputs from more than one deterministic models. To make this simulation study closer to the real life case, we used locations from a dataset studied by Kasibhatla and Chameides (2000); Hogrefe et al. (2001b,a). There, measurements come from monitoring sites ("stations") and modeling output from deterministic models. In fact, we use

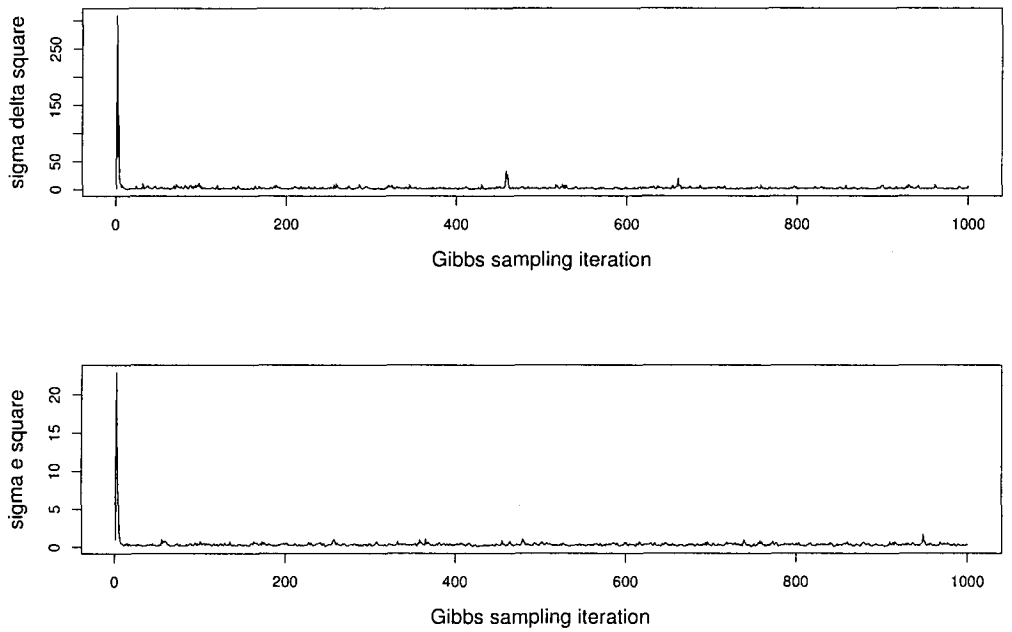

Figure 4.6: MCMC Gibbs samplers of error variance σ_δ^2 and σ_e^2 in the case of 50 grid cells.

only a subset of these stations and grid cells in our simulation, with 50 monitored sites treated as stations and 100 un-monitored sites. For simplicity we assume only two deterministic models in our simulation study yielding output in 20 grid cells and each of them has two sampling points. Figure 4.7 shows the locations of the sites (monitored/unmonitored) and the sampling points within grid cells.

The mean of the true underlying random process $\boldsymbol{Z}$ is a polynomial func-

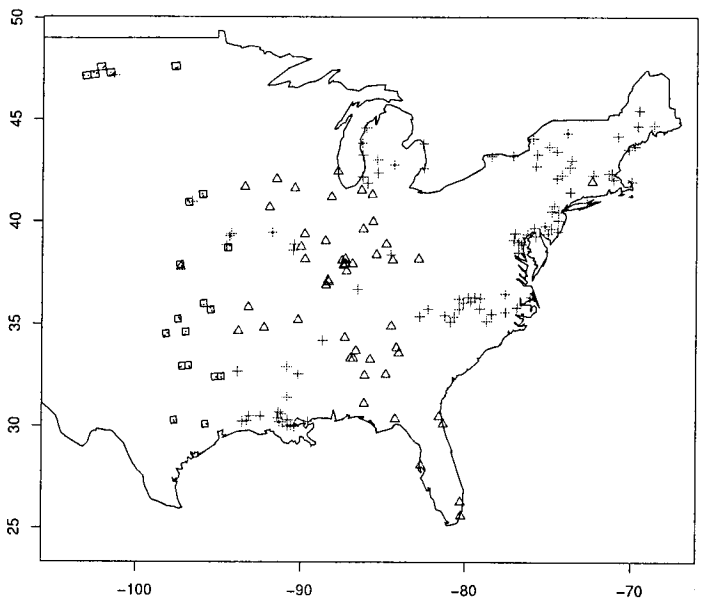

Figure 4.7: Locations of 150 sites (50 monitored, 100 to be predicted) and 20 grid cells. Stations: Δ; sites with sites to be predicted: +; grid cells: rectangles. Each grid cell has two sampling points in it. The x-axis is the longitude in degrees, y-axis, the latitude in degrees.

tion of the coordinates:

$$\mathrm{E}\left[Z(\boldsymbol{s})\right] = \mu(\boldsymbol{s}) = \beta_0 + \beta_1 s_1 + \beta_2 s_2 + \beta_3 s_1^2 + \beta_4 s_2^2 + \beta_5 s_1 s_2,$$

s_1 and s_2 being the coordinates at generated location $\boldsymbol{s}$. This second degree polynomial mean function of the coordinates is one degree higher then the mean function (4.1) used in the previous simulation. This more complicated choice permits us to see how well the BEM model works no matter how

complicated the mean function.

In this simulation, we generate 15 independent datasets. Initially we tried 1000 iterations of the MCMC algorithm but found it failed to converge. Much better results obtained after we extended the number of iterations to 10,000 with a "burn-in" period of 1000. In the case of ensemble deterministic models, there are more parameters in the BEM model. The complexity of BEM model requires a much longer Markov chain to reach convergence.

Simulation Results

Table 4.8 gives the estimated values for the parameters and Table 4.9, the SSPE (sum of squared prediction errors) for the 100 unmonitored stations. These results, suggest the following conclusions.

- Estimates of the additive and multiplicative biases for both deterministic models are quite accurate (Table 4.8).
- The estimate of the *sill* parameter σ is very close to the true value while that of the *range* parameter ρ is not. However, the true value does lie within its 90% credibility interval.
- Estimates of $\sigma^2_{\delta,2}$ is reasonably accurate. However, the estimate of $\sigma^2_{\delta,1}$ exceeds the true value by quite a margin.
- The BEM model gives better predictions than Kriging, measured by SSPE, because the calibration parameters are estimated very well,

meaning in effect, that the model output helps to achieve better prediction.

In summary, the BEM model seems to work quite well when we have more than one deterministic model although the estimate of $\sigma^2_{\delta,1}$ is not very close to the true value.

4.3 Reversible jumps: the stationary case

This simulation involves 20 stations and 50 grid cells. Each grid cell has only one sampling point to reduce the computation burden. All stations (monitoring sites) as well as sampling points are generated uniformly on $[-5, 5] \times [-5, 5]$. The mean function of the underlying true process is

$$\mathrm{E}\left[Z(\boldsymbol{s})\right] = \mu(\boldsymbol{s}) = \beta_0 + \beta_1 s_1 + \beta_2 s_2 + \beta_3 s_1 * s_2$$

with true parameters $\boldsymbol{\beta} = (1.3, 1.2, 0.5, 0.3)$. The true parameters of the exponential covariance function are $\sigma = 1.5$ and $\rho = 5.0$. The calibration parameters are $a = 5.0$ and $b = 2.5$. The variances of the measurement error and model output error are $\sigma_e^2 = 0.25$ and $\sigma_\delta^2 = 0.25$ respectively. We simulate 15 independent dataset in total.

Figure 4.8 shows the MCMC plot and histogram of the dimension k. Table 4.10 shows the estimation results. They lead us to make the following observations.

- The Markov chain for k converges to the true dimension $k = 4$ (Figure 4.8).

- The estimates of $\boldsymbol{\beta}$ are averaged over all the MCMC iterations in which $\boldsymbol{\beta}$ has dimension of 4 and that the estimates are close to the true value of $\boldsymbol{\beta}$.

- The true values of various parameters are within the Bayesian credible intervals except for those of σ_e^2 and σ_δ^2.

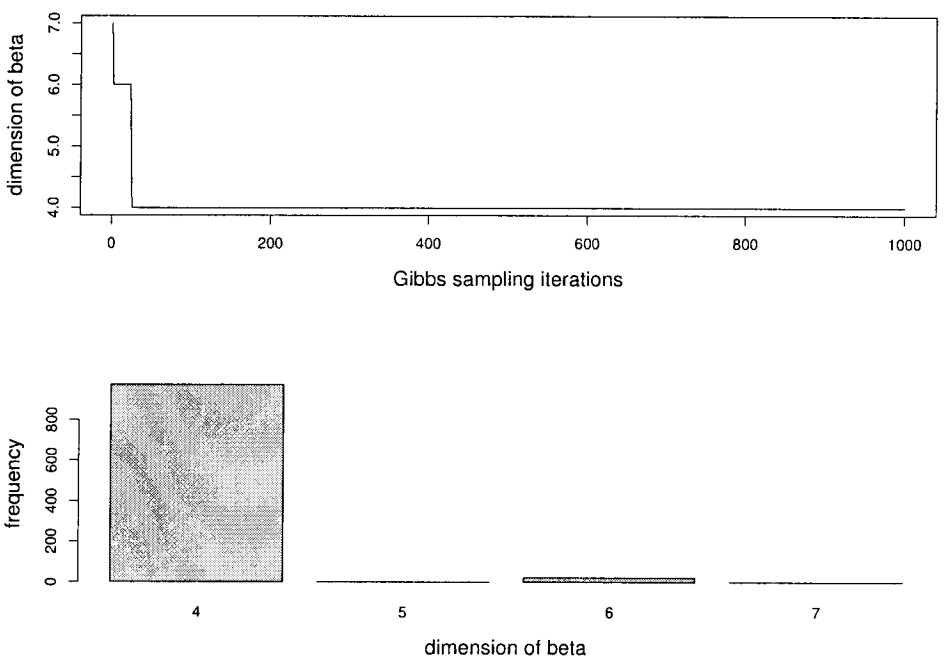

Figure 4.8: The upper plot: the dimension k of the coefficient $\boldsymbol{\beta}$, as a function of the number of MCMC iterations. The lower plot: histogram of the posterior MCMC samples of k.

4.4 Bayesian Melding in Non-stationarity Case

In Equation (2.6) the bandwidth h has to be chosen to represent the non-stationary process by convolution of latent stationary processes. A minimal requirement for h: ensure the positive definiteness of the covariance matrix. Yet if h is small then two locations will have no point, around which stationary processes center, between them, the covariance between these two locations will be zero. So, if h is too small then too many zeros will appear in the covariance matrix of the non-stationary process, resulting in numerical problems when this covariance matrix is inverted to evaluate the likelihood of the non-stationary process.

On the other hand, if h is too big, the covariance between two locations could also be too small because the kernel function is too flat and M in (2.7) is not big enough. In practice M cannot be chosen too big because of the computational burden. This simulation studies the effect of varying h. Primary interests focuses on spatial prediction error as well as on the accuracy of the a and b estimates.

Table 4.11 and 4.12 present the prediction and estimation results. They show that as long as h is not too small, both the SSPE and estimates of a and b are not much different. However, when the chosen h is too big, say, more than twice the true bandwidth used to simulate the data, the SSPE is much bigger. Also, the standard error of estimators of a and b are

much bigger with small h (40% of the true h) than the estimators with a bigger h. If h is chosen too small, say, less than half of the true bandwidth used to simulated the data, even the MCMC algorithm will "crash" because the covariance matrix becomes singular. So as a conservative strategy, we recommend choosing a large h. Large in this can mean half of the distance between two points located at the far ends of the diagonal line across the region. That choice of h is less likely to produce a row or column full of zeros in the spatial covariance matrix, leaving it more numerically stable.

Table 4.4: SSPE (Sum of squared prediction errors) of Kriging and the BM model with different number of grid cells. Since the data is simulated, there is no unit associated to the numbers in this table.

Dataset	Kriging	0cells	2 cells	10 cells	20 cells	30 cells	50 cells
1	116.03	74.11	74.21	75.83	61.75	56.67	71.86
2	104.50	42.16	42.73	40.57	40.60	41.17	39.46
3	223.32	104.80	115.60	107.57	94.54	88.72	78.28
4	119.76	78.73	76.37	73.50	73.48	69.00	63.57
5	182.26	80.53	90.61	72.47	73.20	69.82	62.21
6	135.59	59.63	64.79	52.65	52.60	53.14	51.06
7	150.89	68.86	68.71	70.23	68.27	67.88	67.92
8	127.38	92.14	90.36	94.48	86.12	89.74	79.00
9	115.09	65.81	59.11	62.19	57.87	60.31	56.87
10	121.31	120.22	117.87	106.56	103.36	91.41	82.43
11	107.19	56.38	67.78	57.15	54.68	55.63	51.33
12	172.63	100.56	96.47	96.12	93.62	89.37	86.77
13	140.90	71.37	70.93	69.01	63.82	62.55	56.94
14	194.51	126.80	134.38	134.74	120.24	123.25	104.89
15	181.34	92.89	89.98	83.47	84.37	83.54	78.83
16	257.84	151.52	141.73	129.46	118.79	105.07	95.62
17	141.69	81.64	80.08	77.18	73.68	72.90	57.69
18	170.47	90.68	89.63	91.66	84.61	79.97	73.82
19	113.81	73.24	74.97	74.34	69.21	67.39	66.16
20	99.03	83.48	85.73	82.64	80.42	79.52	73.01
21	249.37	163.09	147.83	156.51	141.12	131.03	105.89
22	138.46	85.70	95.09	81.80	79.31	72.60	64.44
23	138.66	82.25	73.79	81.62	69.68	68.94	61.82
24	110.81	76.74	75.58	79.02	81.83	81.40	69.30
25	110.84	92.90	98.19	88.23	84.67	82.00	76.09
26	162.32	112.48	114.98	110.75	102.47	91.22	85.60
27	203.55	112.85	113.07	116.73	123.22	112.19	105.84
28	138.53	72.54	74.38	73.53	69.50	68.60	66.67
29	117.12	82.46	82.15	101.03	75.22	74.61	69.97
30	152.81	135.98	133.59	122.34	117.14	93.48	72.49
31	157.05	117.43	117.13	107.43	106.66	105.21	92.00
32	140.36	101.53	100.57	93.45	88.29	93.38	88.65
33	142.60	103.28	101.85	104.96	106.54	111.43	107.54
34	102.68	93.99	86.41	88.18	74.78	77.84	67.23
35	153.52	80.53	78.76	75.80	74.19	76.07	70.82
36	171.43	75.52	73.94	70.59	65.64	63.30	52.16
37	105.98	85.41	85.17	79.46	72.99	78.09	66.47
38	144.45	62.95	71.70	63.52	68.56	66.12	60.56
39	257.85	173.02	170.12	163.96	179.29	154.24	132.83
40	158.86	123.88	119.54	109.30	95.47	91.86	104.50
41	145.30	93.46	100.74	101.24	94.68	84.76	65.31
42	128.23	72.57	73.41	73.23	69.72	67.10	65.88
43	193.23	80.03	83.14	82.81	82.26	80.17	75.48
44	130.01	72.40	72.02	74.59	74.97	70.93	67.97
45	112.00	79.34	80.79	72.00	76.52	67.87	81.12
46	168.65	121.84	95.57	91.42	86.39	91.86	76.49
47	213.19	137.96	139.81	122.81	122.18	112.25	93.99
48	100.05	84.13	74.76	76.80	68.59	65.79	63.60
49	136.49	70.64	74.48	66.42	62.05	66.89	54.86
50	113.47	110.61	92.53	89.05	85.96	79.55	73.53
mean	149.47	93.46	92.66	89.41	85.10	81.76	74.74

Table 4.5: The SSPE when the true values of the parameters (a, b, $\boldsymbol{\beta}$, $\boldsymbol{\theta}$) are used. That is, we use the conditional mean of the unavailable measurements based on the available measurements to predict at the 100 unavailable stations. Since the data is simulated, there is no unit associated to the numbers in this table.

Dataset	2 cells	10 cells	20 cells	30 cells	50 cells
1	71.45	67.06	58.71	43.00	43.47
2	40.87	44.22	40.07	41.19	38.79
3	91.50	89.48	83.86	78.88	59.20
4	66.72	62.83	62.69	55.20	56.90
5	82.61	61.58	61.24	61.48	45.55
6	53.37	52.38	49.85	49.16	51.02
7	80.49	78.99	72.17	65.16	62.48
8	75.61	78.63	66.16	67.84	64.87
9	52.14	51.78	50.65	54.46	56.31
10	74.02	71.02	58.54	50.27	44.31
11	59.88	55.76	48.13	46.95	48.71
12	95.59	94.42	75.94	67.41	59.28
13	63.23	61.81	57.52	52.86	55.63
14	91.40	86.96	85.65	69.78	54.15
15	87.79	90.01	85.94	81.63	62.36
16	57.62	56.83	54.90	50.70	44.47
17	66.08	67.89	58.25	60.47	45.96
18	85.65	81.95	61.91	61.83	50.13
19	70.43	72.75	65.97	57.36	48.66
20	77.18	66.47	61.21	56.51	51.70
21	105.28	99.17	103.48	71.71	54.18
22	72.46	75.67	59.15	62.64	53.83
23	53.52	52.24	46.91	48.51	47.59
24	72.67	68.45	64.94	67.81	45.60
25	77.12	68.43	68.92	59.51	60.16
26	89.49	84.16	70.41	68.56	50.87
27	112.37	103.01	72.94	72.33	54.01
28	69.11	58.72	60.26	60.95	56.24
29	63.89	65.79	54.27	52.92	55.52
30	69.43	62.33	59.45	45.74	43.89
31	94.46	89.88	83.92	80.52	74.34
32	80.29	56.97	57.68	47.16	47.38
33	94.49	84.18	86.67	53.36	48.66
34	73.97	62.18	54.61	53.44	45.64
35	77.03	68.71	62.90	65.48	55.35
36	49.02	48.34	45.57	40.47	39.14
37	57.07	56.01	52.06	51.31	54.31
38	54.79	53.17	55.68	58.89	43.27
39	107.22	84.85	67.01	71.63	72.91
40	79.55	75.74	65.22	66.18	57.71
41	93.32	93.24	52.51	54.55	44.87
42	60.07	64.15	59.39	50.81	45.13
43	71.83	74.13	73.27	68.45	51.47
44	71.14	71.26	71.38	64.28	61.59
45	69.84	57.97	54.07	52.08	48.51
46	69.84	61.69	67.31	75.64	70.61
47	78.45	71.75	70.10	56.13	56.87
48	65.29	56.04	45.04	43.28	38.92
49	60.39	55.84	51.19	43.97	42.54
50	80.41	75.18	64.94	58.50	47.76
mean	74.35	69.84	63.21	58.78	52.26

Table 4.6: Coverage probability of the credible interval for the simulation study in Section 4.1, with 20 monitored sites and up to 50 grid cells to predict 100 un-monitored sites. The first column is the nominal coverage probability of the credible interval.

	kriging	0 cell	2 cells	10 cells	20 cells	30 cells	50 cells
95%	37.60%	68.50%	91.98%	94.40%	94.48%	97.23%	95.38%
90%	32.70%	59.68%	84.78%	89.42%	88.22%	86.00%	91.90%
80%	26.32%	49.84%	73.04%	79.58%	78.88%	76.48%	80.76%
70%	21.86%	40.64%	62.04%	70.88%	68.26%	66.28%	71.92%
60%	18.64%	33.50%	52.70%	57.78%	56.24%	55.48%	59.66%
40%	12.22%	22.92%	34.86%	39.00%	38.96%	36.78%	40.72%

Table 4.7: Estimates are based on 20 monitored sites and 50 sampling points. There are 25 sampling points in close proximity to one another. The averages are computed over 50 dataset. True values are $\sigma = 1.50$ and $\rho = 5.00$. Since this table is for the estimates of parameters σ and ρ, there is no unit associated to the numbers in this table.

	averaged posterior mean	averaged posterior sd
σ	0.97	0.25
ρ	4.69	0.76

Table 4.8: Parameter estimates in the simulation study of ensembles of deterministic models in Section 4.2. Columns 3-4 give the averages of posterior means and standard deviations for a_i, b_i, $i = 1, 2$, the calibration parameters of the model output for two deterministic models. $\sigma_{\delta,1}$ and $\sigma_{\delta,2}$ are the variances of two model output error processes. The averages are computed over 15 dataset. Since this table is for the estimates of parameters, there is no unit associated to the numbers in this table. Each of these parameters is a scalar.

Parameters	True value	mean	sd
a_1	5.00	4.94	0.48
b_1	2.50	2.50	0.02
a_2	4.00	4.00	0.68
b_2	3.40	3.39	0.03
σ	1.50	1.51	0.28
ρ	5.00	6.73	0.91
β_0	2.50	1.98	0.59
β_1	2.90	2.92	0.17
β_2	3.20	3.20	0.16
β_3	0.80	0.80	0.03
β_4	1.10	1.11	0.03
β_5	1.30	0.31	0.03
σ_e	0.25	0.33	0.13
$\sigma_{\delta,1}$	0.25	0.24	0.46
$\sigma_{\delta,2}$	0.25	4.51	0.91

Table 4.9: SSPE of the BM model and Kriging in the simulation of multi-deterministic models. We have 50 monitored sites and 100 unmonitored ones. There are 20 grid cells each of them having two sampling points inside. Since the data is simulated, there is no unit associated to the numbers in this table.

Dataset	Kriging	BM
1	42.57	33.13
2	74.07	70.99
3	85.39	59.90
4	56.63	48.59
5	81.75	45.36
6	81.91	73.19
7	56.64	55.89
8	55.31	48.74
9	85.39	56.79
10	72.93	57.10
11	60.41	49.30
12	65.04	59.70
13	97.69	50.43
14	62.23	60.55
15	60.79	41.45
mean	69.25	54.07

Table 4.10: Estimates in the stationary case using reversible jump BM model. Both the "estimates" and "standard error" are averages obtained from 15 independent dataset. Since this table is for the estimates of parameters, there is no unit associated to the numbers in this table. Each of these parameters is a scalar.

Parameters	True value	Estimate	Standard Error
k	4	4.079	0.41
a_1	5.00	5.32	0.52
b_1	2.50	2.45	0.07
σ	1.50	1.10	0.27
ρ	5.00	4.94	0.67
β_0	1.30	0.96	0.74
β_1	1.20	1.06	0.49
β_2	0.50	0.51	0.10
β_3	0.30	0.32	0.04
σ_e^2	0.25	0.39	0.06
σ_δ^2	0.25	4.52	0.83

Table 4.11: SSPEs with different choices of h in non-stationary BM model. Columns 2-6 give the SSPEs for different ratios of the true bandwidth generating the data over the chosen bandwidth in the non-stationary BM model. Since the data is simulated, there is no unit associated to the numbers in this table.

dataset	1.67	1.33	1.00	0.50	0.40
1	7.99	7.74	8.09	8.89	10.73
2	3.39	3.01	2.92	3.26	5.71
3	3.50	3.78	7.40	2.59	4.63
4	4.30	3.60	3.26	3.28	4.24
5	4.11	4.79	4.44	5.18	17.33
6	2.04	4.01	2.18	3.61	2.71
7	3.87	3.87	5.12	4.19	5.58
8	6.15	6.13	7.70	6.40	8.36
9	2.90	3.04	2.49	5.09	18.33
10	4.63	5.94	4.96	7.03	6.68
11	2.49	3.15	3.27	5.11	3.08
12	3.29	2.40	2.56	5.09	7.56
13	4.13	4.40	5.46	5.07	3.50
14	3.27	3.49	3.81	2.90	7.70
15	7.45	7.26	7.87	7.04	6.26
mean	4.23	4.44	4.77	4.98	7.49

Table 4.12: Estimates of a and b for varying values of h in non-stationary BM model. Columns 2-5 give the averages of posterior means and standard deviations. The averages are computed over the 15 datasets generated in the simulation study. The true values are $a = 5.00$ and $b = 2.50$. Since this table is for the estimates of parameters, there is no unit associated to the numbers in this table. Each of these parameters is a scalar.

(true h)/(chosen h)	$\bar{\hat{a}}$	$\bar{sd}$	$\bar{\hat{b}}$	$\bar{sd}$
1.67	5.04	0.33	2.49	0.05
1.33	5.03	0.32	2.49	0.05
1.00	5.01	0.31	2.49	0.05
0.50	5.00	0.30	2.50	0.05
0.40	5.51	1.23	2.72	0.58

Chapter 5

A New Univariate Spatial-Temporal Model

The BM model cannot deal with temporal correlation in the data. To provide an alternative to the BM model, we consider a regression model. The rest of this chapter is organized as follows. Section 5.1 reviews a two-step linear regression model, Section 5.2 presents an ad-hoc extension of the two-step linear regression model. Based on the favorable results seen in Section 5.2, Section 5.3 develops more rigorously a spatial-temporal model in a Bayesian framework. The data analysis results and conclusions are in Section 7.3.

5.1 Two-Step Linear Regression

This section reviews the two-step linear regression model proposed by Guillas et al. (2006). The strong linear correlation between the hourly measurements and model output at each station makes such a model seem natural. Figure 5.1 shows those correlations to be bigger than 0.5 at most stations, pointing to a linear relationship between the measurements and model outputs.

The first model in the two-step linear regression procedure proposed by Guillas et al. (2006) relates the measurements, $\{O(t)\}$, to the model outputs $\{M(t)\}$ by

$$O(t) = c + aM(t) + N_t, \ t = 1, 2, \cdots, T \tag{5.1}$$

with autocorrelated residuals

$$N_t = \rho N_{t-1} + e_t. \tag{5.2}$$

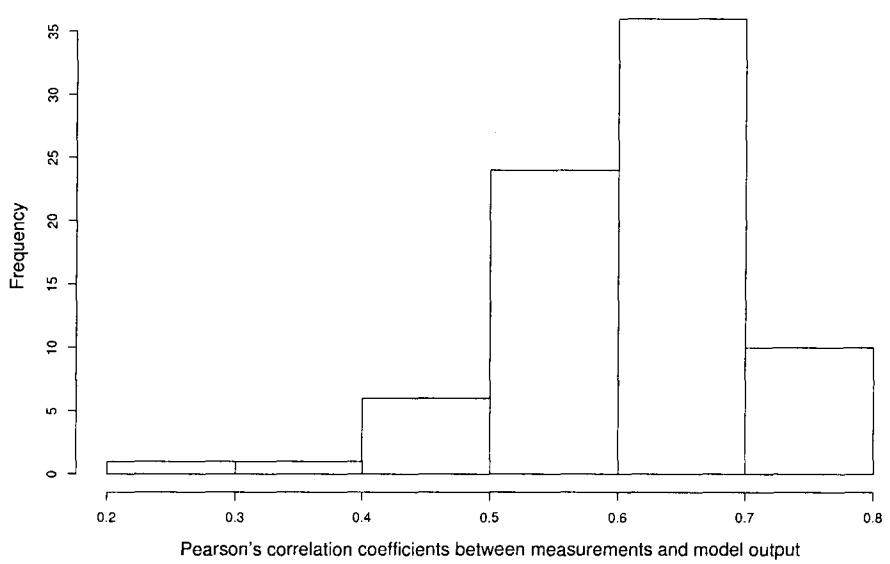

Figure 5.1: Histogram of Pearson's correlation coefficients between hourly measurements and model output for the 78 stations.

The residuals for the autocorrelation model in turn satisfy the following

linear regression model:

$$e_t = \sum_{i=0}^{11} \alpha_i m_i(t) + \epsilon_t, \tag{5.3}$$

$$m_i(t) = \begin{cases} 1 & \text{if t mod 12 = i;} \\ 0 & \text{if t mod 12} \neq \text{i.} \end{cases}$$

$i = 0, \cdots, 11$.

The models (5.1), (5.2) and (5.3) are the same for all stations, while the coefficients are station specific. Both the measurements and model outputs are at the hourly level of temporal resolution and thus they have a strong auto-regressive structure.

$\{X_t\}$ is an AR(p) process if for every t

$$X_t = c + \sum_{i=1}^{p} \rho_i X_{t-i} + \epsilon_t,$$

where $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ independently and identically. The autoregssive parameters are $\rho_1, \cdots, \rho_p$. Many authors omit constant c for simplicity. Some constraints are necessary on the values of the parameters of this model in order that the model remains stationary. For example, processes in the AR(1) model with $|\rho_1| > 1$ are not stationary. Assuming the $\{N_t\}$ have an AR(1) autoregressive structure provides computational simplicity at the possible expense of realism. We also have tried an AR(2) structure for the $\{N_t\}$ but the forecast results do not improve, leading us to use AR(1) throughout

this thesis. Following Guillas et al. (2006), we first fit model (5.1) by using generalized least squares as proposed by Box et al. (1994), $N(t)$ being an AR(1) process. If the model output captures the temporal structure of the measurements very well then model (5.1) would be enough. However, most of the time, the measurements will still have some temporal structure that the model output fails to capture. That is to say, the residuals $\{e_t\}$ will have non-zero means. So, the second step is to fit model (5.3). The covariates in this model are indicator functions of the 12 hours. We also tried indicator functions of the 24 hours, but the forecast results do not improve. So, in favor of fewer parameters, we choose indicator functions of the 12 hours. For the residuals of the AR(1) process $\{N_t\}$, $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ independently and identically.

Models (5.1-5.3) are fitted sequentially and independently. These two models can then be used to forecast the measurements. Suppose at each station, the training data are measurements and model output for times $t = 1$ to T. For times $t = T+1, \cdots, T+T'$, only the model output is available and the goal is to forecast the measurement during that future time period. To do so, models (5.1-5.3) are fitted to obtain estimators $\hat{c}$, $\hat{a}$, $\hat{\rho}$, $\hat{\alpha}_i$, $i = 1, \cdots, 12$. The forecast of $\hat{O}_t$ is then the following

$$\hat{O}_t = \hat{c} + \hat{a}M_t + \hat{N}_t,$$

where

$$\hat{N}_t = \hat{\rho}\hat{N}_{t-1} + \sum_{i=0}^{11} \hat{\alpha}_i m_i(t) \tag{5.4}$$

$t = T+1, \cdots, T+T'$.

Although the two-step regression above is easy to implement, it has two disadvantages. First, it does not include any spatial correlation between stations. We fit models (5.1-5.3) at each station independently. Thus we can only forecast the measurement for stations that have measurements in the past, after estimating the requisite coefficients. In other words, it cannot borrow strength across space to forecast future values of stations without past data by exploiting data from neighbors with a past. Second, the two-step regression procedure only gives point forecasts without any indication of their uncertainties.

To address the first disadvantage of the two-step regression, we first consider a simple but ad-hoc approach which Kriges parameter estimates ($\hat{a}$, $\hat{c}$ and $\hat{\alpha}_i$) across space to get values for them at stations without measurements. After affirming the value of our ad-hoc approach, we address the second disadvantage by extending it to a Bayesian spatial-temporal model. The latter has better Bayesian credentials than the former but that comes at the price of greater complexity. The credible predictive intervals of the forecast and spatial prediction can then be obtained from their MCMC posterior samples at considerable computation cost.

5.2 Two-step linear regression with Kriging

This section presents an ad-hoc approach which is able to not only forecast the measurements but also spatially predict the measurements at those locations without historic measurements. First, we fit models (5.1) and (5.3) at the stations with measurements. Then the estimates of parameters a, c and α_i, $i = 0, \cdots, 11$, are interpolated at locations without measurements by Kriging as described in Section (2.3). We uses the exponential function (2.3) to model the spatial covariance. More complex choices of the spatial covariance function will be considered in future work.

At the locations without measurements but with model output, we have the Kriging interpolated values of the parameters a, c and α_i, $i = 0, \cdots, 11$. Then we can plug these parameters into Formula (5.4) to obtain spatial predictions of the measurements.

5.3 A Bayesian Hierarchical Spatial-Temporal Model

The ad-hoc approach in the previous section is "self-contradictory" because it first fits models (5.1-5.3) to get estimates for $\{\alpha_i\}$ ignoring their spatial correlation. But then it uses Kriging to interpolate the estimated parameter values at the locations without measurements assuming parameters at different stations are spatially correlated.

Another disadvantage of the ad-hoc approach is that it fails to provide measures of the uncertainties associated with the forecasts or predictions. Kassteele et al. (2006) use a similar two-step regression model to interpolate PM10 concentrations over western Europe.

In this paper, the authors first fit a linear regression model with PM10 as the response, modeling output and other explanatory variables as covariates. Then the residuals are interpolated using Kriging. The approach in Kassteele et al. (2006) has the same weakness as the ad-hoc approach presented in the previous section. That is, it assumes the residuals are spatially independent to fit the model by using ordinary least square method and then uses Kriging to interpolate the residuals at other locations. Moreover, the model used in Kassteele et al. (2006) does not include temporal correlation because the data are yearly averages of PM10 concentrations that do not have much temporal correlation. Hence, strictly speaking, both the ad-hoc approach in the previous section and that of Kassteele et al. (2006) do not yield spatial-temporal models. This section presents a Bayesian hierarchical spatial-temporal model that assumes parameters at different stations in models (5.1) and (5.3) are spatially correlated Gaussian processes.

For the Bayesian spatial-temporal model has the following mathematical

form:

$$
\begin{aligned}
&O_{\boldsymbol{s},t} = a_{\boldsymbol{s}} + c_{\boldsymbol{s}} M_{\boldsymbol{s},t} + N_{\boldsymbol{s},t} \\
&N_{\boldsymbol{s},t} = \rho N_{\boldsymbol{s},t-1} + \gamma_{1,\boldsymbol{s}} Z_{1,t} + \gamma_{2,\boldsymbol{s}} Z_{2,t} + \cdots + \gamma_{12,\boldsymbol{s}} Z_{12,t} + \epsilon_{\boldsymbol{s},t} \\
&\boldsymbol{a} = (a_1, \cdots, a_n)^{\mathrm{T}} \sim \mathrm{MVN}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) \\
&\boldsymbol{c} = (c_1, \cdots, c_n)^{\mathrm{T}} \sim \mathrm{MVN}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \\
&\boldsymbol{\gamma}_1 = (\gamma_{1,1}, \cdots, \gamma_{1,n})^{\mathrm{T}} \sim \mathrm{MVN}(\boldsymbol{\mu}_{\gamma_1}, \boldsymbol{\Sigma}_{\gamma_1}) \\
&\cdots \\
&\boldsymbol{\gamma}_{12} = (\gamma_{12,1}, \cdots, \gamma_{12,n})^{\mathrm{T}} \sim \mathrm{MVN}(\boldsymbol{\mu}_{\gamma_{12}}, \boldsymbol{\Sigma}_{\gamma_{12}}) \\
&(\epsilon_{1,t}, \cdots, \epsilon_{n,t})^{\mathrm{T}} \sim \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon) \text{ independently and identically} \\
&\text{times } t = 1, \cdots, T, \text{ and} \\
&\text{sites } \boldsymbol{s} = s_1, \cdots, s_n \\
&\boldsymbol{\mu}_a = (\mu_a, \cdots, \mu_a)^{\mathrm{T}} \\
&\boldsymbol{\mu}_c = (\mu_c, \cdots, \mu_c)^{\mathrm{T}} \\
&\boldsymbol{\Sigma}_a = \sigma_a^2 \exp(-\boldsymbol{D}/\lambda_a) \\
&\boldsymbol{\Sigma}_c = \sigma_c^2 \exp(-\boldsymbol{D}/\lambda_c) \\
&\boldsymbol{\Sigma}_{\gamma_i} = \sigma_{\gamma_i}^2 \exp(-\boldsymbol{D}/\lambda_{\gamma_i}), \ i = 1, \cdots, 12 \\
&\boldsymbol{\mu}_{\gamma_i} = (\mu_{\gamma_i}, \cdots, \mu_{\gamma_i})^{\mathrm{T}}, \ i = 1, \cdots, 12 \\
&\boldsymbol{\Sigma}_\epsilon = \sigma_\epsilon^2 \exp(-\boldsymbol{D}/\lambda_\epsilon),
\end{aligned}
\tag{5.5}
$$

n being the number of monitoring sites, T, the number of hours and D, the

Euclidean distance matrix between stations. Vectors $\boldsymbol{\mu}_a$, $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_{\gamma_i}$ have dimension n. In model (5.5), $\boldsymbol{s}$ denotes location and t, time. The time-related covariates at time t are $Z_{i,t} = 0$ except for $t \bmod 12 = i$ where $Z_{i,t} = 1$, $i = 1, \cdots 12$. These time-related covariates are the same as the indicator functions $m_i(t)$ in (5.3). At each time t, the residuals at different locations are also spatially correlated. The spatial correlation is isotropic and specified by an exponential correlation function. At each location $\boldsymbol{s}$, the residuals $\epsilon_{\boldsymbol{s},t}$ form an AR(1) time series with mean 0.

We make the following assumptions for this Bayesian spatial-temporal model.

- The measurements and model output are linearly related.
- The autoregressive parameter ρ is constant across stations. Section 11.2 has more discussion about this assumption.
- The temporal-spatial correlation is separable. For a process $\{Z(\boldsymbol{s}, t)\}$, its temporal-spatial is separable if the correlation can be written as product of spatial and temporal correlation functions. That is, the temporal correlation does not change over space and the spatial correlation remains the same over time. Section 11.2 discuss future work to release this assumption.
- The spatial correlation between measurements at different stations can be explained by the model output and the spatial correlation between

the coefficients and residuals. Section 7.3 justifies this assumption in our data analysis.

We use the Gibbs sampling algorithm proposed by Gelfand and Smith (1990) to fit the Bayesian spatial-temporal model (5.5). The key to make Gibbs sampling work efficiently is to write model (5.5) in matrix form, that is

$$
\begin{aligned}
&\boldsymbol{O} = \boldsymbol{A}\boldsymbol{a} + \boldsymbol{M}\boldsymbol{c} + \boldsymbol{N} \\
&\boldsymbol{R}\boldsymbol{N} = \boldsymbol{\gamma}_1 \boldsymbol{Z}_1 + \cdots + \boldsymbol{\gamma}_{12} \boldsymbol{Z}_{12} + \boldsymbol{\epsilon} \\
&\boldsymbol{a} \sim \mathrm{MVN}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) \\
&\boldsymbol{c} \sim \mathrm{MVN}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \\
&\boldsymbol{\gamma}_i \sim \mathrm{MVN}(\boldsymbol{\mu}_{\gamma_i}, \boldsymbol{\Sigma}_{\gamma_i}) \quad i = 1, \cdots, 12 \\
&\boldsymbol{\epsilon} \sim \mathrm{MVN}(\boldsymbol{0}, \boldsymbol{I}_{T-1} \otimes \boldsymbol{\Sigma}_\epsilon),
\end{aligned} \tag{5.6}
$$

$\boldsymbol{I}_{T-1}$ being an identity matrix with dimension $(T-1) \times (T-1)$ and $\boldsymbol{\Sigma}_\epsilon$ is the spatial covariance matrix of residuals $\epsilon_{s,t}$ across space at each time t.

The measurement vector is denoted as $\boldsymbol{O} = (O_{1,1}, \cdots, O_{1,n}, \cdots, O_{T,1}, \cdots, O_{T,n})^{\mathrm{T}}$ and the vector $\boldsymbol{N}$, as $\boldsymbol{N} = (N_{1,1}, \cdots, N_{1,n}, \cdots, N_{T,1}, \cdots, N_{T,n})^{\mathrm{T}}$. The measurements are arranged first in the order of locations and then in the order of time. So, the measurements from different locations at the same time are next to each other. That is how we get the Kronecker product structure for the variance matrix of residuals $\boldsymbol{\epsilon}$ in model (5.6). The matrix

form of the model output is

$$\boldsymbol{M} = \begin{pmatrix} M_{1,1} & \dots & 0 \\ \vdots & \ddots & 0 \\ 0 & \dots & M_{1,n} \\ \vdots & \vdots & \vdots \\ M_{t,1} & \dots & 0 \\ \vdots & \ddots & 0 \\ 0 & \dots & M_{t,n} \\ \vdots & \vdots & \vdots \\ M_{T,1} & \dots & 0 \\ \vdots & \ddots & 0 \\ 0 & \dots & M_{T,n} \end{pmatrix}_{nT \times n} .$$

In the matrix $\boldsymbol{M}$, the rows from $(t-1) \times n + 1$ to $t \times n$ form a diagonal matrix of dimension $n \times n$ for $t = 1, \cdots, T$. The elements in each diagonal matrix are the measurements from different locations at time t. The matrix

$\boldsymbol{A}$ has the same structure as $\boldsymbol{M}$. We have

$$\boldsymbol{A} = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & 0 \\ 0 & \dots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \dots & 0 \\ \vdots & \ddots & 0 \\ 0 & \dots & 1 \end{pmatrix}_{nT \times n} .$$

In the matrix $\boldsymbol{A}$, the rows from $(t-1) \times n + 1$ to $t \times n$ form an identity matrix of dimension $n \times n$ for $t = 1, \cdots, T$.

The matrix $\boldsymbol{R}$ has dimensions $n(T-1) \times nT$. At each row i, $i = 1, \cdots, n(T-1)$, the element at column i is $-\rho$ and the element at column $i+n$ is 1. All the other elements in the matrix $\boldsymbol{R}$ are 0.

The matrix $\boldsymbol{Z}_1$ is

$$\boldsymbol{Z}_1 = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & 0 \\ 0 & \dots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \dots & 0 \\ \vdots & \ddots & 0 \\ 0 & \dots & 1 \\ \vdots & \vdots & \vdots \\ 0 & \dots & 0 \\ \vdots & \ddots & 0 \\ 0 & \dots & 0 \end{pmatrix}_{n(T-1)\times n} .$$

This matrix has dimensions $n(T-1) \times n$. The rows from $(t-1) \times n + 1$ to $t \times n$ $(t = 1, \cdots, T)$ form an identity matrix of dimension $n \times n$ only for $t \bmod 12 = 1$. All the other elements in the matrix $\boldsymbol{R}$ are 0. The matrices $\boldsymbol{Z}_i$ $(i = 2 \cdots 12)$ have the same dimensions as matrix $\boldsymbol{Z}_1$. For matrix $\boldsymbol{Z}_i$ $(i = 2 \cdots 12)$, the rows from $(t-1) \times n + 1$ to $t \times n$ $(t = 1, \cdots, T)$ form an identity matrix of dimension $n \times n$ only for $t \bmod 12 = i$. All the other elements in the matrix $\boldsymbol{R}$ are 0.

Let $\boldsymbol{N} = \boldsymbol{O} - \boldsymbol{Aa} - \boldsymbol{Mc}$ and $\boldsymbol{F} = \sum_{i=1}^{12} \gamma_i \boldsymbol{Z}_i$. Then the density of

$\boldsymbol{\epsilon} = \boldsymbol{RN}$ is the following:

$$p(\boldsymbol{O}|\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{R}, \Sigma_\epsilon) \propto |\boldsymbol{\Sigma}_\epsilon|^{-\frac{T-1}{2}}$$
$$\exp\left(-\frac{1}{2}\left[\boldsymbol{R}(\boldsymbol{O} - \boldsymbol{Aa} - \boldsymbol{Mc}) - \boldsymbol{F}\right]^{\mathrm{T}}(\boldsymbol{I}_{T-1} \otimes \boldsymbol{\Sigma}_\epsilon)^{-1}\left[\boldsymbol{R}(\boldsymbol{O} - \boldsymbol{Aa} - \boldsymbol{Mc}) - \boldsymbol{F}\right]\right). \tag{5.7}$$

The inferences about the parameters, forecasts and predictions are based on their Markov Chain Monte Carlo (MCMC) samples generated by the Gibbs sampling algorithm. That algorithm as implemented is iterative. First, we choose some arbitrary initial values for all the parameters. Then, in each iteration of the algorithm, the parameters are updated by generating a random sample from their conditional distributions given the data and other parameter values from the previous iteration. For computational efficiency, each of the parameter vectors $\boldsymbol{a}$, $\boldsymbol{c}$ and $\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_1$ is updated as a block. For completeness, we list the conditional distributions for each of the parameters from (5.8) to (5.16) .

The prior and full conditional distributions of $\boldsymbol{a}$ are

$$p(\boldsymbol{a}|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) \sim \mathrm{MVN}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$$
$$p(\boldsymbol{a}|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a, \boldsymbol{M}, \boldsymbol{A}, \boldsymbol{R}, \boldsymbol{\Sigma}_\epsilon) \sim \mathrm{MVN}(\boldsymbol{\mu}_a^{'}, \boldsymbol{\Sigma}_a^{'})$$
$$\boldsymbol{\Sigma}_a^{'} = \left[\boldsymbol{\Sigma}_a^{-1} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{I}_{T-1} \otimes \boldsymbol{\Sigma}_\epsilon)^{-1}\boldsymbol{RA}\right]^{-1}$$
$$\boldsymbol{\mu}_a^{'} = \boldsymbol{\Sigma}_a^{'}\left[\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{I}_{T-1} \otimes \boldsymbol{\Sigma}_\epsilon)^{-1}\boldsymbol{R}(\boldsymbol{O} - \boldsymbol{Mc})\right.$$
$$\left.-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{I}_{T-1} \otimes \boldsymbol{\Sigma}_\epsilon)^{-1}\boldsymbol{F} + \boldsymbol{\Sigma}_a^{-1}\boldsymbol{\mu}_a\right]. \tag{5.8}$$

The prior and full conditional distributions of $\boldsymbol{c}$ are

$$
\begin{aligned}
&p(\boldsymbol{c}|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \sim \mathrm{MVN}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \\
&p(\boldsymbol{c}|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c, \boldsymbol{M}, \boldsymbol{M}, \boldsymbol{R}, \boldsymbol{\Sigma}_\epsilon) \sim \mathrm{MVN}(\boldsymbol{\mu}_c^{'}, \boldsymbol{\Sigma}_c^{'}) \\
&\boldsymbol{\Sigma}_c^{'} = \left[\boldsymbol{\Sigma}_c^{-1} + \boldsymbol{M}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{I}_{T-1} \otimes \boldsymbol{\Sigma}_\epsilon)^{-1}\boldsymbol{R}\boldsymbol{M}\right]^{-1} \\
&\boldsymbol{\mu}_c^{'} = \boldsymbol{\Sigma}_c^{'}\left[\boldsymbol{M}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{I}_{T-1} \otimes \boldsymbol{\Sigma}_\epsilon)^{-1}\boldsymbol{R}(\boldsymbol{O} - \boldsymbol{A}\boldsymbol{a})\right. \\
&\left.-\boldsymbol{M}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{I}_{T-1} \otimes \boldsymbol{\Sigma}_\epsilon)^{-1}\boldsymbol{F} + \boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c\right]. \qquad (5.9)
\end{aligned}
$$

The prior and full conditional distributions of $\boldsymbol{\gamma}_i$, $i = 1, \cdots, 12$, are

$$
\begin{aligned}
&p(\boldsymbol{\gamma}_i|\boldsymbol{\mu}_{\gamma_i}, \boldsymbol{\Sigma}_{\gamma_i}) \sim \mathrm{MVN}(\boldsymbol{\mu}_{\gamma_i}, \boldsymbol{\Sigma}_{\gamma_i}) \\
&p(\boldsymbol{\gamma}_i|\boldsymbol{\mu}_{\gamma_i}, \boldsymbol{\Sigma}_{\gamma_i}, \boldsymbol{M}, \boldsymbol{A}, \boldsymbol{R}, \boldsymbol{\sigma}_\epsilon) \sim \mathrm{MVN}(\boldsymbol{\mu}_{\gamma_i}^{'}, \boldsymbol{\Sigma}_{\gamma_i}^{'}) \\
&\boldsymbol{\Sigma}_{\gamma_i}^{'} = \left[\boldsymbol{\Sigma}_{\gamma_i}^{-1} + \boldsymbol{Z}_i^{\mathrm{T}}(\boldsymbol{I}_{T-1} \otimes \boldsymbol{\Sigma}_\epsilon)^{-1}\boldsymbol{Z}_i\right]^{-1} \\
&\boldsymbol{\mu}_{\gamma_i}^{'} = \boldsymbol{\Sigma}_{\gamma_i}^{'}\left[\boldsymbol{Z}_i^{\mathrm{T}}(\boldsymbol{I}_{T-1} \otimes \boldsymbol{\Sigma}_\epsilon)^{-1}\left(\boldsymbol{R}(\boldsymbol{O} - \boldsymbol{A}\boldsymbol{a} - \boldsymbol{M}\boldsymbol{c})\right.\right. \\
&\left.\left.-F + \boldsymbol{\gamma}_i\boldsymbol{Z}_i\right) + \boldsymbol{\Sigma}_{\gamma_i}^{-1}\boldsymbol{\mu}_{\gamma_i}\right]. \qquad (5.10)
\end{aligned}
$$

The prior and full conditional distributions of μ_a are both univariate

normal distributions with the following forms:

$$
\begin{aligned}
&p(\mu_a|a_0, f_0) \sim N(a_0, f_0) \\
&p(\mu_a|a_0, f_0, \boldsymbol{\Sigma}_a, \boldsymbol{a}) \sim N(a_0^{'}, f_0^{'}) \\
&f_0^{'} = \left[f_0^{-1} + 2\mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}_a^{-1}\mathbf{1}\right]^{-1} \\
&a_0^{'} = f_0^{'}\left[\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma}_a^{-1}\mathbf{1} + \mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}_a^{-1}\boldsymbol{a}\right],
\end{aligned} \tag{5.11}
$$

$\mathbf{1}$ being a vector of dimension n with all the elements are 1, f_0 and $f_0^{'}$, the prior and conditional variance of μ_a. The full conditional distributions of μ_c, μ_{γ_i} $(i = 1, \cdots, 12)$ are almost same as (5.11) only replacing $\boldsymbol{a}$, Σ_a with $\boldsymbol{c}$, Σ_c and $\boldsymbol{\gamma}_i$, Σ_{γ_i} correspondingly.

The prior and full conditional distributions of σ_a^2 are both inverse gamma distributions with the following forms:

$$
\begin{aligned}
&p(\sigma_a^2|\alpha, \beta) \propto (\sigma_a^2)^{-\alpha-1}\exp\left(-\frac{1}{\beta\sigma_a^2}\right) \\
&p(\sigma_a^2|\alpha, \beta, \boldsymbol{a}, \lambda_a, \boldsymbol{\mu}_a) \propto (\sigma_a^2)^{-\alpha^{'}-1}\exp\left(-\frac{1}{\beta^{'}\sigma_a^2}\right) \\
&\alpha^{'} = \alpha + \frac{n}{2} \\
&\beta^{'} = \left[1/\beta + \frac{1}{2}(\boldsymbol{a} - \boldsymbol{\mu}_a)^{\mathrm{T}}(\exp(-\boldsymbol{D}/\lambda_a))^{-1}(\boldsymbol{a} - \boldsymbol{\mu}_a)\right]^{-1}.
\end{aligned} \tag{5.12}
$$

We cannot find a conjugate prior for the parameter λ_a, so its conditional

distribution does not have a closed form. However, we have

$$p(\lambda_a|\sigma_a^2, \boldsymbol{\mu}_a, \boldsymbol{a}) \propto |\exp(-\boldsymbol{D}/\lambda_a)|^{-1}$$
$$\exp\{-\frac{1}{2}(\boldsymbol{a}-\boldsymbol{\mu}_a)^{\mathrm{T}}(\exp(-\boldsymbol{D}/\lambda_a))^{-1}(\boldsymbol{a}-\boldsymbol{\mu}_a)\}. \tag{5.13}$$

To obtain a random sample from the above distribution of λ_a, we have to use the Metropolis-Hasting algorithm proposed by Metropolis et al. (1953) and Hastings (1970). The conditional distributions of σ_c^2, λ_c and $\sigma_{\gamma_i}^2$,λ_{γ_i} $(i = 1, \cdots, 12)$ are very similar to σ_a^2 and λ_a. We only need to replace $\boldsymbol{a}$, $\boldsymbol{\mu}_a$ with $\boldsymbol{c}$, $\boldsymbol{\mu}_c$ and $\boldsymbol{\gamma}_i$,$\boldsymbol{\mu}_{\gamma_i}$, correspondingly in (5.12) and (5.13).

The prior and conditional distributions of σ_ϵ^2 are both inverse gamma distributions with the following forms:

$$p(\sigma_\epsilon^2|\alpha, \beta) \propto (\sigma_\epsilon^2)^{-\alpha-1}\exp\left(-\frac{1}{\beta\sigma_\epsilon^2}\right)$$
$$p(\sigma_\epsilon^2|\alpha, \beta, \boldsymbol{\epsilon}, \lambda_\epsilon, \boldsymbol{\mu}_\epsilon) \propto (\sigma_\epsilon^2)^{-\alpha'-1}\exp\left(-\frac{1}{\beta'\sigma_\epsilon^2}\right)$$
$$\alpha' = \alpha + \frac{n(T-1)}{2}$$
$$\beta' = \left[1/\beta + \frac{1}{2}\sum_{t=2}^{T}\left(\tilde{\boldsymbol{\epsilon}}_t^{\mathrm{T}}(\exp(-\boldsymbol{D}/\lambda_\epsilon))^{-1}\tilde{\boldsymbol{\epsilon}}_t\right)\right]^{-1}, \tag{5.14}$$

$\tilde{\boldsymbol{\epsilon}}_t$ being the residuals vector at different locations at time t. As with λ_a, we use Metropolis-Hastings algorithm to obtain a random sample of λ_ϵ. We

have

$$p(\lambda_\epsilon|\sigma_\epsilon^2, \boldsymbol{\epsilon}) \propto |\exp(-\boldsymbol{D}/\lambda_\epsilon)|^{-1}$$
$$\exp\{-\frac{1}{2}\sum_{t=1}^{T}\left(\tilde{\boldsymbol{\epsilon}}_t{}^{\mathrm{T}}(\exp(-\boldsymbol{D}/\lambda_\epsilon))^{-1}\tilde{\boldsymbol{\epsilon}}_t\right)\}. \tag{5.15}$$

To update ρ, its ordinary least square (OLS) estimate given other parameters is used instead of a random sample from its conditional distribution. We do this in part for computational simplicity. Finding a conjugate prior for ρ is very difficult while finding its ordinary least square estimation is very easy. The main reason is that the variance of the conditional distribution of ρ is small enough to be treated as zero because of the large sample size $n \times (T-1)$. The OLS estimate of ρ is

$$\hat{\rho} = \frac{\sum_{s=1}^{n}\sum_{t=2}^{T}(\tilde{N}_{s,t} - \sum_{j=1}^{12}\tilde{Z}_{j,t}\gamma_{j,s})\tilde{N}_{s,t-1}}{\sum_{s=1}^{n}\sum_{t=2}\tilde{N}_{s,t-1}^2}, \tag{5.16}$$

where $\tilde{N}$ is the matrix form of $\boldsymbol{N}$ in model (5.6) and $\tilde{Z}_{j,t} = 0$ except for $t \bmod 12 = j$ where $\tilde{Z}_{j,t} = 1$.

At each iteration in Gibbs sampling, the forecast or prediction is obtained in an iterative way. The forecast or prediction is

$$\hat{\boldsymbol{O}}_t = a_s + c_s\boldsymbol{M}_t + \boldsymbol{N}_t$$
$$\boldsymbol{N}_t = \rho\boldsymbol{N}_{t-1} + \sum_{i=1}^{12}\gamma_{i,s}\boldsymbol{Z}_{i,t} + \boldsymbol{\epsilon}_t$$
$$\boldsymbol{\epsilon}_t \text{ is a random sample generated from } \mathrm{MVN}(\mathbf{0}, \Sigma_\epsilon), \tag{5.17}$$

$\boldsymbol{M}_t$ being the vector of measurements from all the stations at time t, $bd\epsilon_t$, the residuals at time t, $\hat{\boldsymbol{O}}_t$ the forecast or prediction at time t. At those stations where the historic measurements are available up to time T, $\hat{\boldsymbol{O}}_t$, $t = T+1, \cdots, T+T'$ are the forecasts. At those stations where there is no historic measurement, $\hat{\boldsymbol{O}}_t$, $t = 1, \cdots, T$ are the predictions. The parameters at those stations without measurements can be spatially predicted by the parameter values at stations with measurement. Finally, the credible interval of the forecast or prediction is obtained by taking the quantiles of all their MCMC samples.

Chapter 6

Multivariate Spatial-Temporal Model

Both the BM model in Chapter 2 and the Bayesian spatial-temporal model in Chapter 5 are univariate models. That is, the responses in these models are univariate variables. However, in practice, the monitoring stations can record the concentration levels of more than one pollutant. It is also possible for the deterministic models to generate multivariate outputs. For example, Grimit and Mass (2002) describe the University of Washington mesoscale short-range ensemble forecasting system over the Pacific Northwest. This system can generate temperature and sea surface level pressure at the same time.

The multivariate regression problem has been studied extensively by Brown and Zidek (1982); van der Merwe and Zidek (1980); Brown and Zidek (1980); Cripps et al. (2003), etc. In geo-statistics, a well known tool for interpolating a multivariate spatial process is the so-called co-Kriging approach described by Wackernagel (1998). Co-Kriging is a multivariate extension of the Kriging approach. However, like Kriging, co-Kriging is purely a spatial model which

cannot handle the temporal correlation in a fundamental way. The more recent literature of multivariate pollutant prediction includes Mardia and Goodall (1993); Brown et al. (1994); Le and Zidek (1997); Sun et al. (1998); Zidek et al. (2000); Schmidt and Gelfand (2003); Gelfand et al. (2004). The related chapters in the books by Le and Zidek (2006); Banerjee et al. (2003) provide excellent summaries of various multivariate models used in the spatial statistics. However, many of multivariate models proposed before assume the responses are temporally independent. In some cases, this assumption is valid because the data has been monthly or yearly averaged so there is not much temporal correlation left. In other cases, some preliminary analysis has been done to remove the temporal correlation before using the multivariate models. This chapter presents a multivariate model which is an extension of the univariate spatial-temporal model proposed in Chapter 5. It incorporates both spatial and temporal correlation. This multivariate model is built in a Bayesian framework in order to evaluate the uncertainty of the forecasts and spatial predictions. The rest of this chapter is organized as follows. Section 6.1 presents the details of the multivariate spatial-temporal model. Section 6.2 presents the MCMC algorithm to fit this multivariate spatial-temporal model.

6.1 Multivariate Spatial-Temporal Model

This section presents the model specifications and full conditional distributions of all the parameters needed to implement the MCMC algorithm. At time t and station $\boldsymbol{s}$, the model has the following form:

$$
\begin{aligned}
&\boldsymbol{O}_{\boldsymbol{s},t} = \boldsymbol{\beta}_{\boldsymbol{s}} \boldsymbol{M}_{\boldsymbol{s},t} + \boldsymbol{N}_{\boldsymbol{s},t} \\
&\boldsymbol{N}_{\boldsymbol{s},t} = \boldsymbol{\rho} \boldsymbol{N}_{s,t-1} + \boldsymbol{\gamma}_{\boldsymbol{s}} \boldsymbol{Z}_{\boldsymbol{s},t} + \boldsymbol{\epsilon}_{\boldsymbol{s},t} \\
&t = 1, \cdots T \\
&\boldsymbol{s} = 1, \cdots n \\
&\boldsymbol{\epsilon}_{\boldsymbol{s},t} \sim \mathrm{MVN}(0, \boldsymbol{\Sigma}_{\epsilon}) \text{ independently and identically,}
\end{aligned} \tag{6.1}
$$

$\boldsymbol{O}_{\boldsymbol{s},t} : q \times 1$ being the measurements vector of q pollutants and $\boldsymbol{M}_{\boldsymbol{s},t} : (p + 1) \times 1$, the vector of intercept term and p model outputs for q pollutants. It is not necessary that $p = q$. The covariate $\boldsymbol{Z}_{s,t}$ is a vector of dimension m, which can include m other variables helping to forecast or predict the measurements. For example, $\boldsymbol{Z}$ can include temperature which is helpful to forecast or predict the ozone concentration levels. The matrix $\boldsymbol{\Sigma}_{\epsilon}$ is the covariance matrix of q pollutants, the so-called co-regionalization matrix in geo-statistics. The auto-regression coefficient matrix ρ is of dimensions $q \times q$.

The coefficients $\boldsymbol{\beta}_{\boldsymbol{s}}$ and $\boldsymbol{\gamma}_{\boldsymbol{s}}$ at each location $\boldsymbol{s}$ are matrices of dimensions $q \times (p + 1)$ and $q \times m$ respectively. Each element of the matrix $\boldsymbol{\beta}_s$ and $\boldsymbol{\gamma}_s$

are spatially correlated across stations. That is, we have

$$
\begin{aligned}
&(\boldsymbol{\beta}_{i,j,1}, \cdots, \boldsymbol{\beta}_{i,j,n})^{\mathrm{T}} \sim \mathrm{MVN}(\boldsymbol{\mu}_{\beta,i,j}, \boldsymbol{\Sigma}_{\beta,i,j}) \\
&\boldsymbol{\Sigma}_{\beta,i,j} = \sigma^2_{\beta,i,j} \exp(-\boldsymbol{D}/\lambda_{\beta,i,j}) \\
&i = 1, \cdots, q, \ j = 0, \cdots, p \\
&(\boldsymbol{\gamma}_{i,j,1}, \cdots, \boldsymbol{\gamma}_{i,j,n})^{\mathrm{T}} \sim \mathrm{MVN}(\boldsymbol{\mu}_{\gamma,i,j}, \boldsymbol{\Sigma}_{\gamma,i,j}) \\
&\boldsymbol{\Sigma}_{\gamma,i,j} = \sigma^2_{\gamma,i,j} \exp(-\boldsymbol{D}/\lambda_{\gamma,i,j}) \\
&i = 1, \cdots, q, \ j = 1, \cdots, m.
\end{aligned} \tag{6.2}
$$

Spatial processes $\{\beta_{i,j,s}\}$ $(i = 1, \cdots, q,\ j = 0, \cdots, p)$ and $\{\gamma_{i,j,s}\}$ $(i = 1, \cdots, q,$ $j = 1, \cdots, m)$ are assumed to have exponential spatial correlation functions. However, these spatial processes, $\{\beta_{i,j,s}\}$ and $\{\gamma_{i,j,s}\}$), have independent prior distributions. We use the Gibbs sampling algorithm to fit models (6.1) and (6.2). We write model (6.1) in a matrix form to make the MCMC algorithm more efficient. The matrix form is the following:

$$
\begin{aligned}
&\boldsymbol{O} = \boldsymbol{M}\boldsymbol{\beta} + \boldsymbol{N} \\
&\boldsymbol{R}\boldsymbol{N} = \boldsymbol{\gamma}\boldsymbol{Z} + \boldsymbol{\epsilon} \\
&\boldsymbol{\epsilon} \sim \mathrm{MVN}(\boldsymbol{0}, \boldsymbol{I} \otimes \boldsymbol{\Sigma}_{\epsilon}),
\end{aligned} \tag{6.3}
$$

the dimension of identity matrix $\boldsymbol{I}$ being $n(T-1) \times n(T-1)$, the measurements vector $\boldsymbol{O}$, formed by stacking vectors $\boldsymbol{O}_t$ $(t = 1, \cdots, T)$. In other words, $\boldsymbol{O}_t = (O_{1,1,t}, \cdots, O_{q,1,t}, \cdots, O_{1,k,t}, \cdots, O_{q,k,t}, O_{1,n,t}, \cdots, O_{q,n,t})^{\mathrm{T}}$, which

includes the measurements of q pollutants for n stations at time t. In model (6.3), we have the coefficients vectors

$\boldsymbol{\beta} = (\beta_{1,0,1}, \cdots, \beta_{1,0,n}, \cdots, \beta_{i,j,1}, \cdots, \beta_{i,j,n}, \cdots, \beta_{q,p,1}, \cdots, \beta_{q,p,n})^{\mathrm{T}}$ and

$\boldsymbol{\gamma} = (\gamma_{1,1,1}, \cdots, \gamma_{1,1,n}, \cdots, \gamma_{i,j,1}, \cdots, \gamma_{i,j,n}, \cdots, \gamma_{q,m,1}, \cdots, \gamma_{q,m,n})^{\mathrm{T}}$. The coefficient $\beta_{i,j,k}$ ($i = 1, \cdots, q$, $j = 1, \cdots, p$, $k = 1, \cdots, n$) is the effect of j-th model output on the i-th measurement at station k and $\beta_{i,0,k}$ is the intercept of the i-th measurement at station k. In principle, we can arrange the elements of vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in arbitrary orders. The reason to have the specific ordering above is that we can easily write the covariance matrix of the coefficients β and γ in a block diagonal form shown later in the next section. In model (6.3), the model output $\boldsymbol{M}$ is matrix of dimension $nqT \times nq(p+1)$. The covariate matrix $\boldsymbol{Z}$ is of dimensions $nqT \times nqm$. The matrix of measurement $\boldsymbol{M}$ is formed by stacking the matrices $\boldsymbol{M}_t$, $t = 1, \cdots, T$ by rows. Each matrix $\boldsymbol{M}_t$ is of dimensions $nq \times nq(p+1)$ and it is formed by stacking the matrices $\boldsymbol{m}_i$, $i = 1, \cdots, n$ by rows, $\boldsymbol{m}_i$ being a diagonal block matrix of dimensions $q \times nq(p+1)$. The row vectors on the diagonal of matrix $\boldsymbol{m}_1$ is $(M_{0,1,t}, 0 \cdots, 0, M_{1,1,t}, 0 \cdots 0, \cdots, M_{p,1,t}, 0 \cdots, 0)$, $M_{0,1,t} = 1$ corresponding to the intercept, $M_{1,1,t}, \cdots, M_{p,1,t}$, p covariates at station 1 at time t. The matrix $\boldsymbol{Z}$ has the same structure as $\boldsymbol{M}$ except that its dimensions are of $nq(T-1) \times nqm$.

The density function of $\boldsymbol{O}$ is the following:

$$p(\boldsymbol{O}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{R},\boldsymbol{\Sigma}_\epsilon) \propto |\boldsymbol{I}\otimes\boldsymbol{\Sigma}_\epsilon|^{-\frac{1}{2}}$$
$$\exp\left(-\frac{1}{2}\left[\boldsymbol{R}(\boldsymbol{O}-\boldsymbol{M}\boldsymbol{\beta})-\boldsymbol{Z}\boldsymbol{\gamma}\right]^{\mathrm{T}}(\boldsymbol{I}\otimes\boldsymbol{\Sigma}^{-1})\left[\boldsymbol{R}(\boldsymbol{O}-\boldsymbol{M}\boldsymbol{\beta})-\boldsymbol{Z}\boldsymbol{\gamma}\right]\right) \quad (6.4)$$

6.2 MCMC Algorithm

To use the Gibbs sampling algorithm, we first need to derive the full conditional distributions of all the parameters. The likelihood function of the parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\rho}$ is (6.4). The prior and full conditional distributions of $\boldsymbol{\beta}$ are the following:

$$p(\boldsymbol{\beta}|\boldsymbol{\mu}_\beta,\boldsymbol{\Sigma}_\beta) \sim \mathrm{MVN}(\boldsymbol{\mu}_\beta,\boldsymbol{\Sigma}_\beta)$$
$$p(\boldsymbol{\beta}|\boldsymbol{\mu}_\beta,\boldsymbol{\Sigma}_\beta,\boldsymbol{R},\boldsymbol{O},\boldsymbol{Z},\boldsymbol{\Sigma}_\epsilon,\boldsymbol{\gamma}) \sim \mathrm{MVN}(\boldsymbol{\mu}_\beta^{'},\boldsymbol{\Sigma}_\beta^{'})$$
$$\boldsymbol{\Sigma}_\beta^{'} = \left[\boldsymbol{\Sigma}_\beta^{-1}+\boldsymbol{M}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{I}\otimes\boldsymbol{\Sigma}_\epsilon^{-1})\boldsymbol{R}\boldsymbol{M}\right]^{-1}$$
$$\boldsymbol{\mu}_\beta^{'} = \boldsymbol{\Sigma}_\beta^{'}\left[\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta+\boldsymbol{M}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{I}\otimes\boldsymbol{\Sigma}_\epsilon^{-1})(\boldsymbol{R}\boldsymbol{O}-\boldsymbol{Z}\boldsymbol{\gamma})\right].$$

The prior mean $\boldsymbol{\mu}_\beta$ is formed by stacking vectors $\boldsymbol{\mu}_{\beta,i,j} = (\mu_{\beta,i,j},\cdots,\mu_{\beta,i,j})^{\mathrm{T}}$, $i = 1,\cdots,q$ and $j = 0,\cdots,p$. The n elements in vector $\boldsymbol{\mu}_{\beta,i,j}$ are identical because we assume the prior mean of $\beta_{i,j}$ is the same at each station. The prior variance of $\boldsymbol{\beta}$ is $\boldsymbol{\Sigma}_\beta = \mathrm{diag}(\boldsymbol{\Sigma}_{\beta,1,0},\cdots,\boldsymbol{\Sigma}_{\beta,i,j},\cdots,\boldsymbol{\Sigma}_{\beta,q,p})$. Each matrix $\boldsymbol{\Sigma}_{\beta,i,j}$ is the spatial covariance matrix of vector $(\boldsymbol{\beta}_{i,j,1},\cdots,\boldsymbol{\beta}_{i,j,n})$, that is, the coefficients $\beta_{i,j}$ at different stations.

Because the spatial processes of coefficients $\beta_{i,j}$ are independent of each other so $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ is a diagonal block matrix of $p \times (p+1)$ blocks.

The prior and full conditional distributions of $\boldsymbol{\gamma}$ are the following:

$$
\begin{aligned}
&p(\boldsymbol{\gamma}|\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}) \sim \mathrm{MVN}(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}) \\
&p(\boldsymbol{\gamma}|\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}, \boldsymbol{R}, \boldsymbol{O}, \boldsymbol{Z}, \boldsymbol{\Sigma}_{\epsilon}, \boldsymbol{\gamma}) \sim \mathrm{MVN}(\boldsymbol{\mu}_{\boldsymbol{\gamma}}^{'}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{'}) \\
&\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{'} = \left[\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} + \boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{I} \otimes \boldsymbol{\Sigma}_{\epsilon}^{-1})\boldsymbol{Z}\right]^{-1} \\
&\boldsymbol{\mu}_{\boldsymbol{\gamma}}^{'} = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{'}\left[\Sigma_{\boldsymbol{\gamma}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\gamma}} + \boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{I} \otimes \boldsymbol{\Sigma}_{\epsilon}^{-1})\boldsymbol{R}(\boldsymbol{O} - \boldsymbol{M}\boldsymbol{\beta})\right].
\end{aligned}
$$

The prior mean and variance of $\boldsymbol{\mu}_{\boldsymbol{\gamma}}$ have the same structure as those of $\boldsymbol{\mu}_{\boldsymbol{\beta}}$. The only difference is that the dimension of the coefficient matrix $\boldsymbol{\gamma}_k$ at each station k is of $q \times m$ but the dimensions of $\boldsymbol{\beta}_k$ is $q \times (p+1)$.

A $q \times q$ positive definite matrix $\boldsymbol{W}$ is said to have a Wishart distribution with v degrees of freedom if its density function is of the form

$$
\begin{aligned}
f(\boldsymbol{W}) &= \left[2^{vq/2}\Gamma_q(v/2)\right]^{-1}|\boldsymbol{A}|^{-v/2}|\boldsymbol{W}|^{v-q-1}\exp\left[-\mathrm{tr}(\boldsymbol{A}^{-1}\boldsymbol{W})/2\right] \\
\Gamma_q(t) &= \pi^{q(q-1)/4}\Pi_{i=1}^{q}\Gamma[t-(i-1)/2] \\
\Gamma(x) &= \int_0^{\infty} t^{x-1}e^{-t}dt. \qquad (6.5)
\end{aligned}
$$

for any positive $q \times q$ matrix $\boldsymbol{A}$, $\mathrm{tr}(\boldsymbol{A}^{-1}\boldsymbol{W})$ being the trace of matrix $\boldsymbol{A}^{-1}\boldsymbol{W}$.

Let $\boldsymbol{W} = \boldsymbol{\Sigma}^{-1}$ stands for the precision matrix in the density function

(6.4). The prior and full conditional distributions of $\boldsymbol{W}$ are

$$
\begin{aligned}
&p(\boldsymbol{W}) \sim \text{Wishart}(\boldsymbol{A}, v) \\
&p(\boldsymbol{W}|\tilde{\boldsymbol{e}}) \sim \text{Wishart}(\boldsymbol{A}', v') \\
&\boldsymbol{A}' = \left(\boldsymbol{A}^{-1} + \tilde{\boldsymbol{e}}^{\mathrm{T}}\tilde{\boldsymbol{e}}\right)^{-1} \\
&v' = v + n(T-1). \qquad (6.6)
\end{aligned}
$$

The $n(T-1) \times q$ matrix $\tilde{\boldsymbol{e}}$ is formed by stacking row vectors $\boldsymbol{e}_{s,t}$, $t = 2, \cdots, T$ and $s = 1, \cdots, n$. Each residuals vector $\boldsymbol{e}_{s,t}$ is of dimension q.

To update the auto-regression coefficient matrix $\boldsymbol{\rho}$, we use ordinary least squares estimation instead of sampling from its full conditional distribution. In Gibbs sampling algorithm, iteratively update each parameter block by its full conditional distributions derived as above in this section. At each iteration in Gibbs sampling, the forecast or prediction is obtained in an iterative way. The forecast or prediction is

$$
\begin{aligned}
&\hat{\boldsymbol{O}}_{s,t} = \boldsymbol{\beta}_s \boldsymbol{M}_{s,t} + \boldsymbol{N}_{s,t} \\
&\boldsymbol{N}_{s,t} = \boldsymbol{\rho}\boldsymbol{N}_{s,t-1} + \boldsymbol{\epsilon}_{s,t} \\
&\boldsymbol{\epsilon}_{s,t} \text{ is a random sample generated from MVN}(\boldsymbol{0}, \boldsymbol{\Sigma}_\epsilon). \qquad (6.7)
\end{aligned}
$$

At those stations where the historic measurements are available up to time T, $\hat{\boldsymbol{O}}_{s,t}$, $t = T+1, \cdots, T+T'$ are the forecasts. At those stations where there is no historic measurements, $\hat{\boldsymbol{O}}_{s,t}$, $t = 1, \cdots, T$ are the predictions.

The parameters at those stations without measurements can be spatially predicted by the parameter values at stations with measurements. Finally, the credible interval of the forecast or prediction is obtained by taking the quantiles of all their MCMC samples. Section 7.4 presents an application of the Bayesian multivariate spatial-temporal model and demonstrates its use.

Chapter 7

Forecasting and Spatial Prediction

One of the important purposes of deterministic models is to predict the phenomena they represent. This chapter first presents spatial prediction results for BM described in Chapter 2. Then it continues to explore the making of such predictions and forecasts by the spatial-temporal model proposed in Chapter 5. In the rest of this chapter, we use the term "prediction" to refer to spatial prediction, while "forecast" refers to future responses.

7.1 Bayesian Melding Model Prediction

The data comes from two sources: regional surface ozone concentration measurements and model outputs from a deterministic model AQM (air quality model), a non-hydrostatic version of the MAQSIP (Multiscale Air Quality Simulation Platform) model. This AQM system has been described in detail by Wheeler and Houyoux (1998). The AQM model output is based on grid cells with resolution 6×6 km^2. The measurements are from the Air Quality

System (AQS) monitoring network. Both the measurements and model outputs are hourly concentrations starting from May 15 to September 11, 1995 over a 120-day period. The dataset represents 375 monitoring stations in the AQS network and 307 grid cells in the AQM output. The measurements and model outputs are based on different supports, that is, measurements are based on point values while model outputs are grid cell averages over space. In addition, the model outputs are based on the UTC (Coordinated Universal Time) time standard, the measurements on local time. Ignoring this time difference would result in poor correlation between measurements and model outputs. The next section shows the necessity to adjust this discrepancy.

This dataset has been analyzed by Kasibhatla and Chameides (2000), Hogrefe et al. (2001b) as well as Hogrefe et al. (2001a). The first paper compares the Pearson correlation coefficients between quantiles (10th, 25th, 50th, 75th and 90th) of the measurements and AQM model outputs for daily average data. The latter two papers also analyze the correlation between measurements and model outputs, but after decomposing the hourly time series into sub-series on different time scales. They reach similar conclusions, that the AQM model output represents the measurement better at longer time scales.

However in estimating the correlation between measurements and model output, the above analysis ignores the temporal and spatial correlation in the data, thus leaving some uncertainty about their validity. Moreover correlation analysis does not help in the assessment of the model calibration

(additive and multiplicative) since the correlation is by definition invariant under the transformation of measurement scales. Finally our interest lies not merely in the degree of linear association between measurements and model outputs measured by the correlation, but rather in predicting the ozone level at unmonitored locations. These factors motivate a new data analysis using the BM.

Preliminary data analysis

The previous section points out that the measurements are based on local time while the model outputs are based on GMT (Greenwich Mean Time). So we adjusted the model outputs to the local time. The dataset has 119 days after the adjustment. Figure 7.1 shows the linear relationship between the measurements and model outputs.

The measurement series, unlike that for the model outputs, have missing values. For example, all the measurements from station 550730005 (in Wisconsin state) are missing. To deal with the missing values, we first choose those stations that have no more than 100 hours of missing measurements. Second, we use the 24 hour mean to fill in the missing values. For example, if the missing values occur at 1000, then we use the average of the available values at 1000 every day to fill in these missing values. After adjusting for different time standards and ignoring the stations with more than 100 missing measurements, we have measurements at 81 stations and model outputs

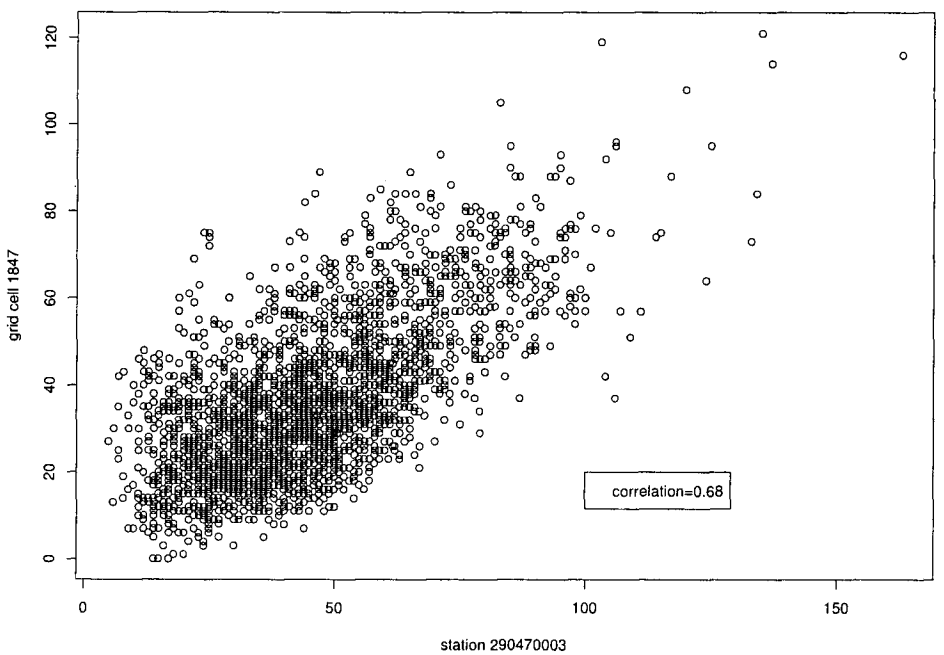

Figure 7.1: Scatter plots and Pearson correlations between measurements of station #290470003 and model outputs for cell #1847. The y-axis stands for the model outputs of grid cell #1847 and the x-axis stands for the measurements at station #290470003. The station #290470003 is inside grid cell #1847. Observe that the correlation between measurements and model outputs is 0.7.

on 375 grid cells of 2856 hours (119 days). In these 375 grid cells, there are 78 grid cells which contain one and only one station. To enable us understand better the role of model-to-measurement correlation in spatial prediction, from now on the data always will focus on the 78 grid cells with 78 stations inside the grid cells. Although the measurements and model outputs are available during the 119 day period, we only focus on the 30 days of July, when the ozone concentration is at a high level due to high temperatures.

The ozone concentration level at night and early in the morning is much lower than during the day. Figure 7.2 shows the 24 side-by-side hourly box plots of measured ozone levels at station #10731005 and simulated ozone levels at grid cell #3529. The station #10731005 is inside grid cell #3529. Both the observed and simulated ozone level are at a peak during the 8 hours from 1000 to 1700. Figure 7.3 shows the histograms of the model-to-measurement correlation at all 24 hours and at the 8 hours. Based on Figures 7.2 and 7.3, we focus on the analysis of 8-hour measurements and model outputs, since otherwise the two data sources are quite dissimilar. Moreover, little interest obtains in the hours outside that period. From now on, the 8-hour average measurements or model outputs is also referred to as the "daily average". We apply BM to analyze the hourly, daily and weekly average ozone concentration levels for the selected 8-hour time period. Kasibhatla and Chameides (2000); Fiore et al. (2003, 2004) focus their analysis on similar daily average data. In particular, Fiore et al. (2003) use the empirical orthogonal function (EOF) method, a principal component analysis (PCA), to compare the first

two principal components of the measurement and model outputs.

The simulated ozone levels at the 78 grid cells and measurements at 48 stations are used to fit the BM model (2.9). The measurements at the remaining 30 stations are used as validation data. These 30 stations are called "un-monitoring" sites from now on. Figure 7.4 maps the grid cells and monitoring/un-monitoring sites. For locations $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, we use the Euclidean distance of their longitude and latitude to define the distance between them as the following:

$$\text{dist}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \sqrt{(s_{1,1} - s_{2,1})^2 + (s_{1,2} - s_{2,2})^2}, \tag{7.1}$$

$s_{1,1}$ and $s_{2,1}$ being the longitudes of location $\boldsymbol{s}_1$, $s_{1,2}$ and $s_{2,2}$, the latitudes of location $\boldsymbol{s}_2$. In fact, the more accurate formula to calculate the distance in km between two locations is the following:

$$\text{dist}(\boldsymbol{s}_1, \boldsymbol{s}_2) = 111.21 \times \sqrt{(s_{1,1} - s_{2,1})^2 + \cos^2(s_{1,2}/57.3)(s_{1,2} - s_{2,2})^2}. \tag{7.2}$$

The distances calculated from Equations (7.1) and (7.2) are not proportional because of the cosine term in Equation (7.2). However, the relative errors of the distances between the 78 stations we have is small. Figure 7.5 shows that about 90% of the relative errors are within 10%. Another reason for us to use the Equation (7.1) to define the distance between stations is for the usage of Kriging in R package "geoR" by Ribeiro and Diggle (2001). That package define the distance by using Equations (7.1). To compare the results

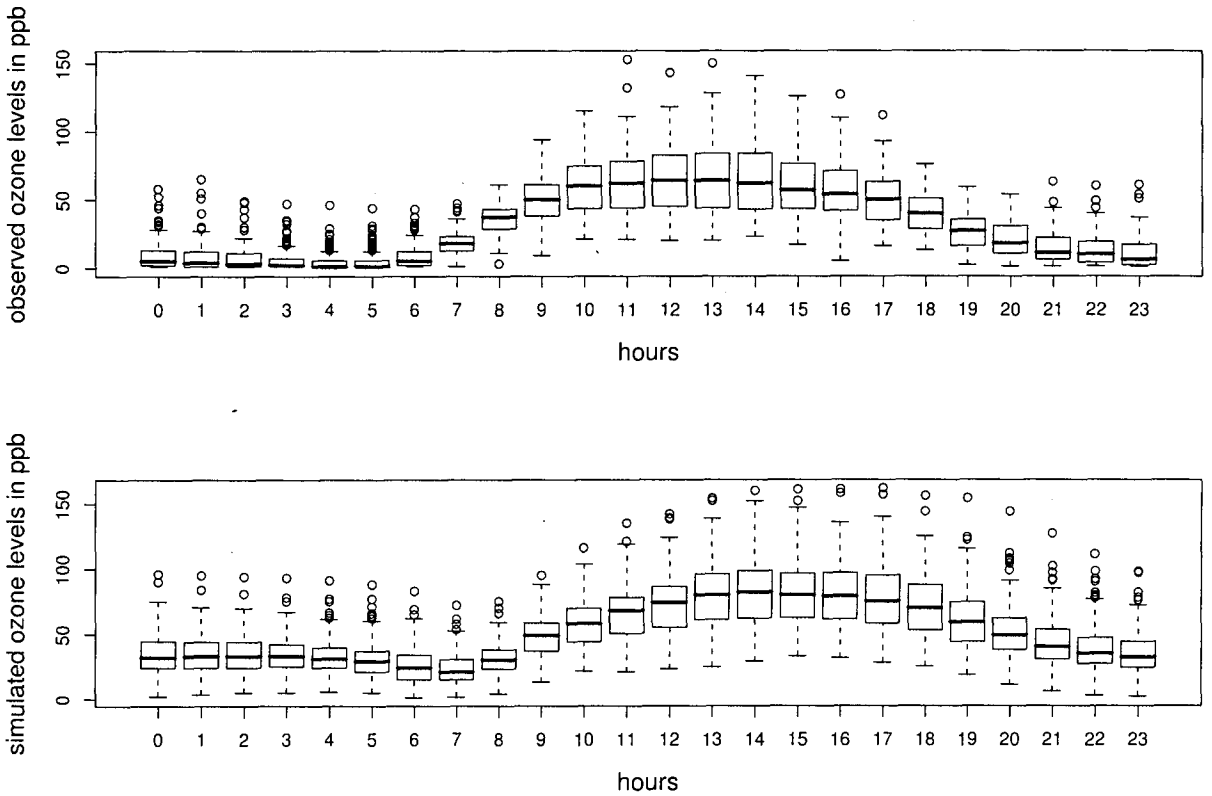

Figure 7.2: The upper panel is the 24 side-by-side hourly box plots of measurements at station #10731005. The lower panel is the 24 side-by-side hourly box plots of model outputs (simulated ozone levels) in grid cell #3529. The station #10731005 lies inside grid cell #3529.

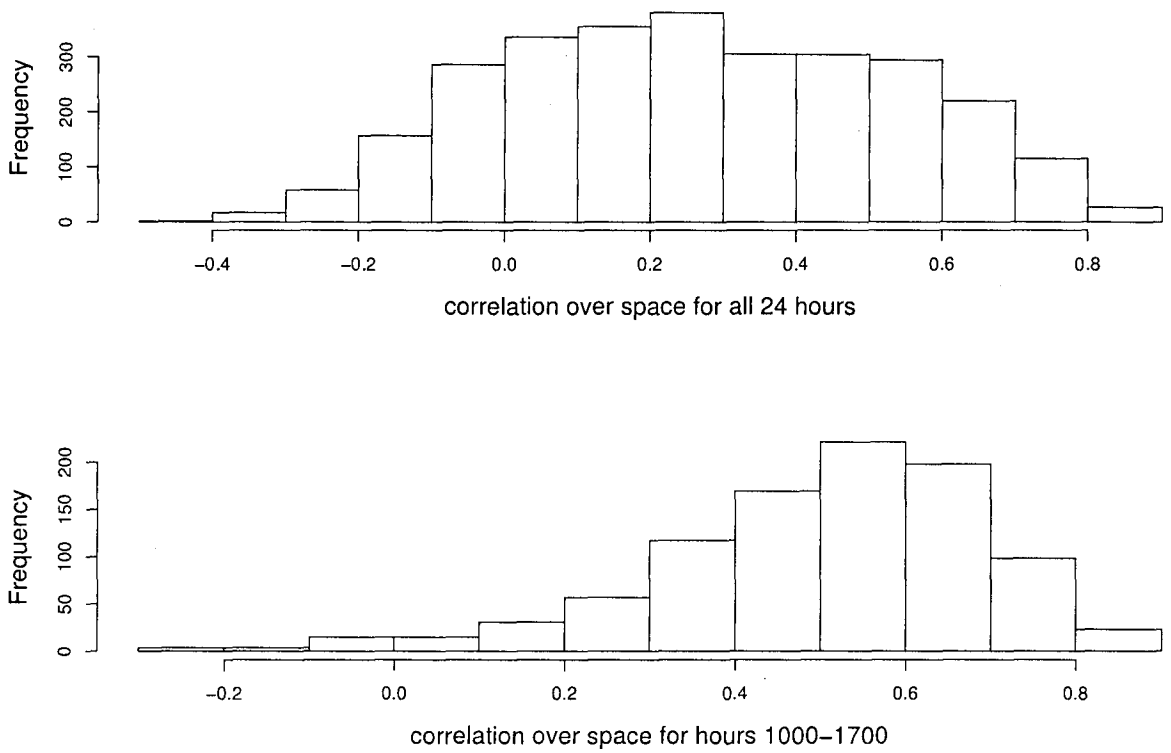

Figure 7.3: Histograms of correlation over space at each hour. The upper histogram is for all the 24 hours. The lower histogram is for the 8-hour day time. Notice the markedly larger model-to-measurement correlations during the hours 1000-1700.

of Kriging and other approaches, we stick to that definition of the distance from now on.

We compare the results of spatial prediction between two competing approaches: BM and Kriging. In fact, we have two different versions of Kriging: Kriging using measurements and Kriging using model outputs. The RMSPE (root mean square prediction error) measures the predictive performance. At time t, which could be hour, day or week, we define the RMSPE by

$$\text{RMSPE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - \hat{O}_i)^2}, \tag{7.3}$$

n being the number of un-monitoring sites to be predicted, O_i, the measurement at station i and $\hat{O}_i$, the prediction.

Analysis of Hourly Data

For the 30 days from July 1 to July 30, 1995, we have 240 hourly measurements and model outputs at the daily 8-hour time period (1000-1700) selected for analysis. We apply the BM model and Kriging to each hour separately. For brevity, we list the average RMSPE for the BM model and Kriging in Table 7.1. The averages are computed over the 240 hours. Figure 7.6 shows the plot of the RMSPE difference between the BM model and Kriging using only measurements versus the correlation between measurements and model outputs over space. We offer the following observations about the analysis

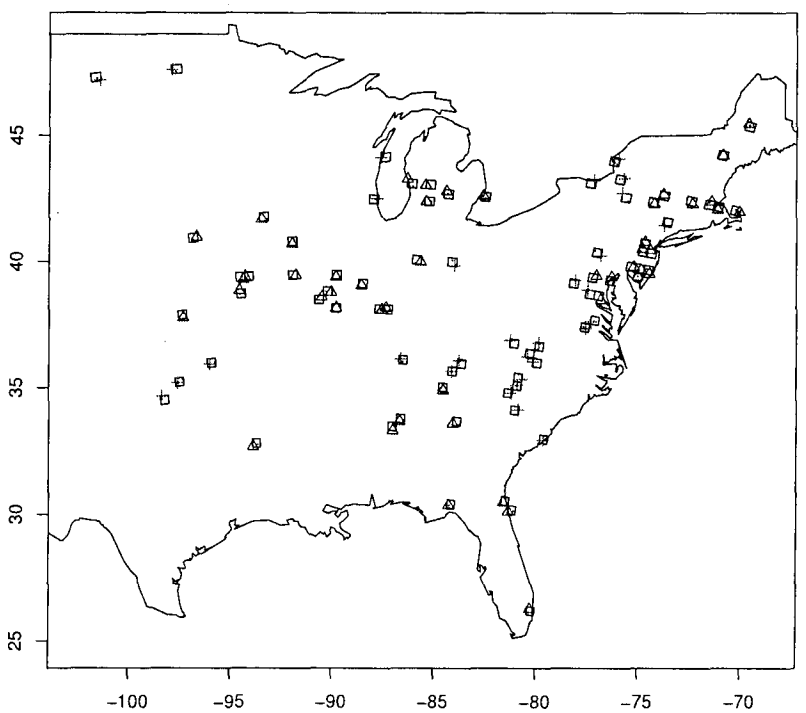

Figure 7.4: Locations of 48 available stations, 78 grid cells and 30 unavailable stations. Δ: available stations. Rectangle: grid cells. $+$: unavailable stations to be predicted. The x-axis is the longitude in degrees, y-axis, the latitude in degrees.

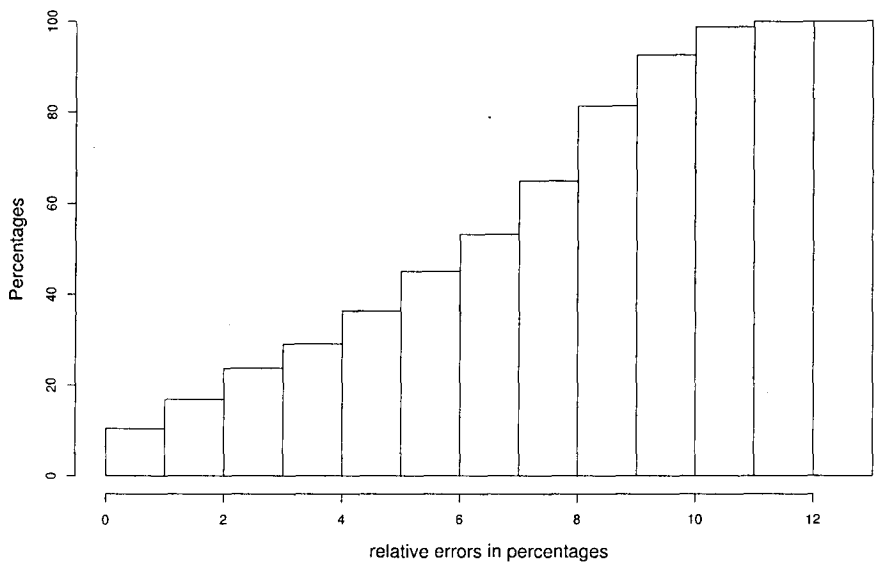

Figure 7.5: Cumulative plots of the relative errors in percentages between the distance using formula 7.1 and formula 7.2.

of hourly data.

- Table 7.1 shows that on average, the BM predictor has the smallest RMSPE among all the competitive prediction approaches.

- The BM predictor seems marginally better than Kriging using only the measurements. The RMSPE for the BM predictor is 15.82 and for Kriging using measurements, 16.36. As shown in the next section, the BM predictor does much better than that version of Kriging in the analysis of daily average data.

- Kriging using model outputs has the biggest RMSPE, pointing to the desirability of calibrating the model outputs.

- Figure 7.6 shows that in general, for spatial prediction, BM more likely outperforms Kriging using measurements when the correlation between measurements and model outputs is big. Note there is one hour at which the BM prediction is much worse than Kriging. The reason is that the estimation of variance by BM is huge at that hour. This also reveals some instability of BM with the hourly data, which maybe one of the reasons that Fuentes and Raftery (2005) use the weekly average data in their analysis.

- The coverage probability of BM's 90% predictive interval is 86.72%, which is fairly good.

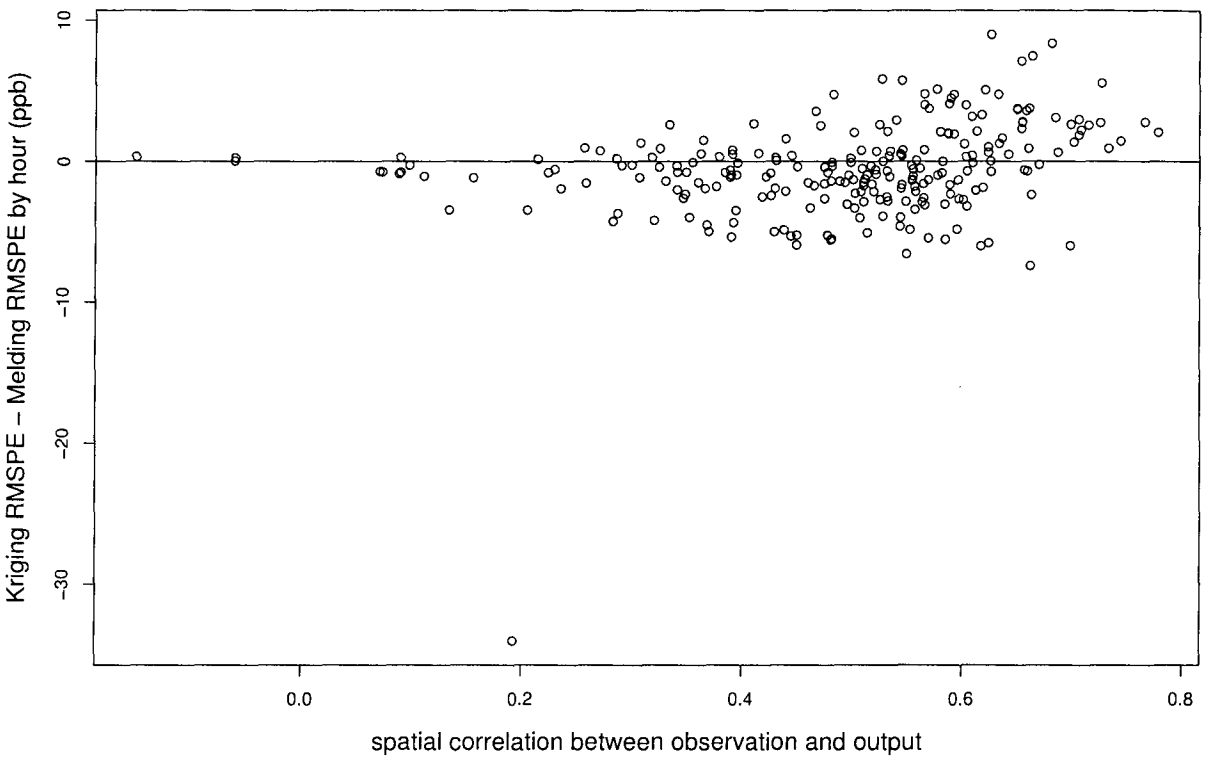

Figure 7.6: Hourly RMSPE (root mean square prediction error) difference between Kriging using measurements and BM versus the correlation over space between measurements and model outputs. Points above the horizontal line in the plot represent victories for the BM model over Kriging.

Analysis of 8-hour Daily Average Data

This section presents our analysis of daily 8-hour (1000-1700) measurements and model outputs averages. We would make the following observations about our analysis of daily average data.

- Table 7.2 shows that for daily averages, the BM model has the smallest RMSPE among all competitive prediction approaches. The RMSPE of the BM model is smaller than Kriging using measurements on 18 out of the 30 days. The average RMSPE for Kriging using model outputs is the biggest, showing the need to calibrate the model outputs.

- Figure 7.7 confirms the intuitively plausible result that in general the prediction performance of the BM model improves as the correlation between measurements and model outputs increases. When the correlation exceeds 0.6, the BM model performs substantially better than Kriging on most days.

- In implementing BM, we assume the additive calibration parameter $a(\boldsymbol{s})$ is a polynomial function of the coordinates at location $\boldsymbol{s}$, that is $a(\boldsymbol{s}) = f(\boldsymbol{s})\boldsymbol{\beta}_a$. We use the reversible jump MCMC to choose the degree of the polynomial function. The reversible jump MCMC can sample the dimension of $\boldsymbol{\beta}_a$ from its posterior distribution. Section 3.6 describes this reversible jump MCMC in detail. For all 30 days, the posterior distribution of $\boldsymbol{\beta}_a$ ranges between 1 to 3. So, we can

assume $a(\boldsymbol{s})$ is a linear function of the coordinates at location $\boldsymbol{s}$, that is, $a(\boldsymbol{s}) = a_0 + a_1 s_1 + a_2 s_2$, s_1 and s_2 being the longitude and latitude in degrees at location $\boldsymbol{s}$. By assuming a is a linear function coordinates at location $\boldsymbol{s}$, we get a smaller RMSPE than by assuming a is a constant across the space.

- In BM (2.9), we assume that the multiplicative calibration parameter b is constant across stations. The more realistic this assumption, the better will be the BM predictor's performance against Kriging. In the dataset, each station is located in a grid cell, so the model outputs for that grid cell can be treated as the model output at the station inside the grid cell. Thus, at each grid cell $\boldsymbol{B}$, we can plausibly estimate b by

$$\hat{b}(\boldsymbol{B}) = \frac{\tilde{Z}(\boldsymbol{B}) - \hat{a}(\boldsymbol{B})}{\hat{Z}(\boldsymbol{B})}, \tag{7.4}$$

where $\hat{Z}(B) = \int_B \hat{Z}(\boldsymbol{s})d\boldsymbol{s}$ is the integral of $\{\hat{Z}(\boldsymbol{s})\}$ over the grid cell B. Because we only have one sampling point in each grid cell, this integral is just $\hat{Z}(\boldsymbol{s})$, that is, the measurement at the station inside grid cell B. After obtaining $\hat{b}(B)$ from the above formula, we can compute the sampling variance of $\{\hat{b}(B)\}$ and the absolute mean difference of $\{\hat{b}(B)\}$ between grid cells having measured or unmeasured stations. Figure 7.8 and Figure 7.9 show plots of (Kriging RMSPE - BM RMSPE) versus $\hat{\mathrm{Var}}(\hat{b})$ and the absolute mean difference for $\hat{b}$, respectively, for the 30 days in July. From these plots we can see that the BM prediction is

better than Kriging when $\hat{\text{Var}}(\hat{b})$ and the absolute mean difference of $\hat{b}$ are small. This finding is expected because the BM prediction is better when the model assumptions are more justified.

- The coverage probability of the 90% credible for the BM prediction is 87.33%, which is fairly good.

Analysis of weekly Average Data

Fuentes and Raftery (2005) analyze weekly average SO_2 measurements and model outputs in USA. The available/unavailable stations and grid cells used in the analysis of weekly average data is the same as in the analysis of daily average data. However, with averages over longer time scales, the predictions of both the BM model and Kriging are expected to improve because of better normality of the data distribution and smaller variation of the average data over the longer time scale. The improvement of the BM prediction also lies in the better performance of deterministic model for the longer time scale as noted by Kasibhatla and Chameides (2000); Hogrefe et al. (2001b,a). Table 7.3 presents that the RMSPE of the BM model , Kriging with only measurements, Kriging with measurement plus model outputs without calibration, Kriging with only model outputs without calibration. We have the following findings from the analysis of weekly average data.

- Table 7.3 shows that the BM model achieves the smallest RMSPE in

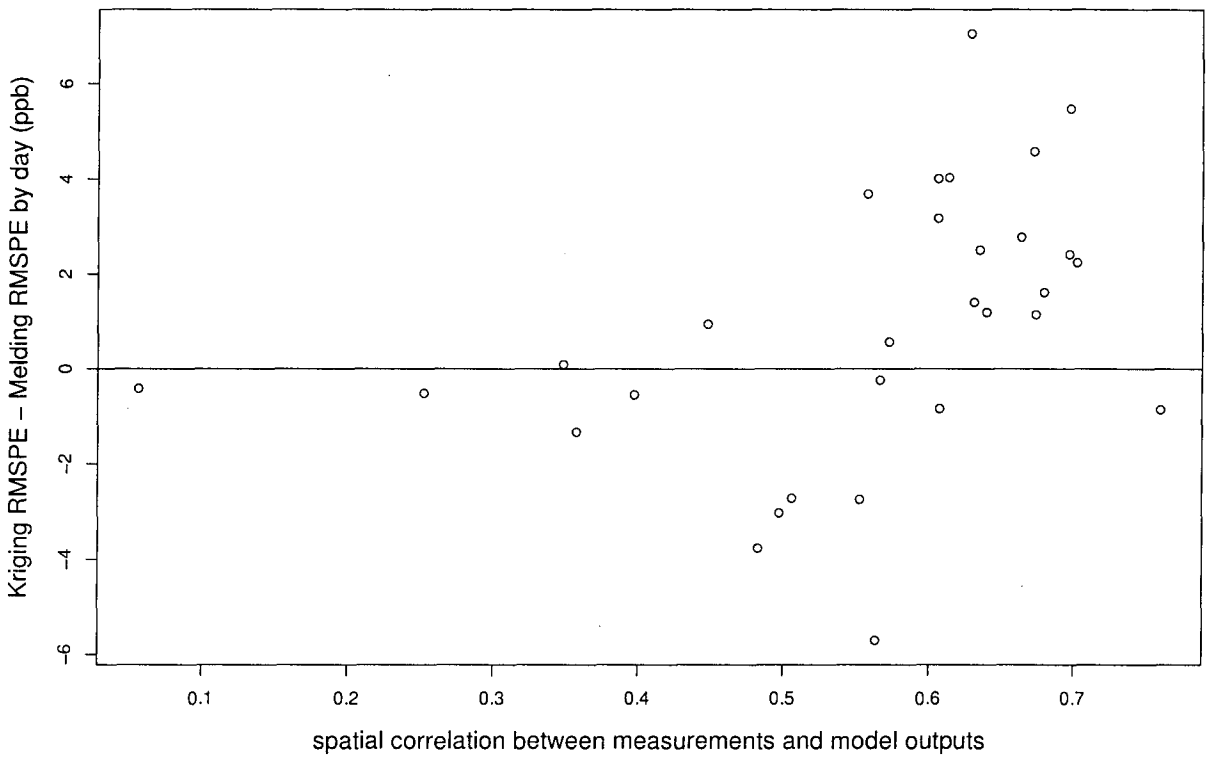

Figure 7.7: RMSPE difference between Kriging prediction with measurements and the BM prediction versus the correlation over space between measurements and model outputs. Points above the plotted horizontal line represented victory for the BM model. Notice the supremacy of the BM model when correlation exceeds 0.6.

Table 7.1: Average RMSPE of hourly ozone predictions at 30 stations. Column 1: RMSPE for BM model. Column 2: RMSPE for Kriging using measurements. Column 3: RMSPE for Kriging using model outputs. The unit of the numbers in this table is ppbv.

BM	Kriging 1	Kriging 2
15.82	16.36	16.91

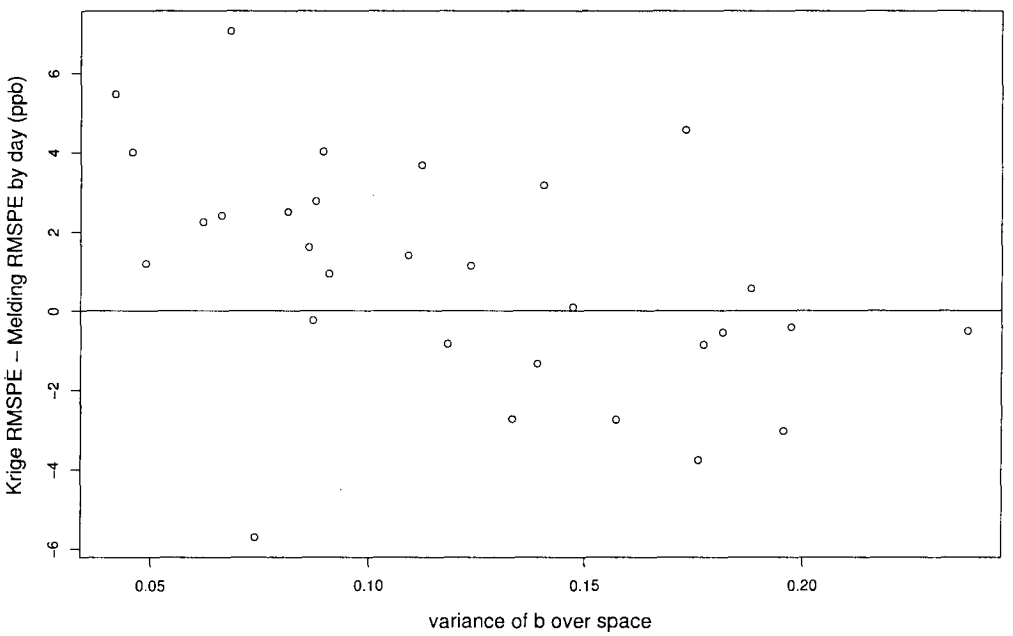

Figure 7.8: Daily RMSPE difference between Kriging with measurements and the BM model versus the variance of $\hat{b}$ (defined by formula (7.4)) across space. Points above the plotted horizontal line represented victory for the BM model. Notice the supremacy of the BM model when the variance of $\hat{b}$ over space is small.

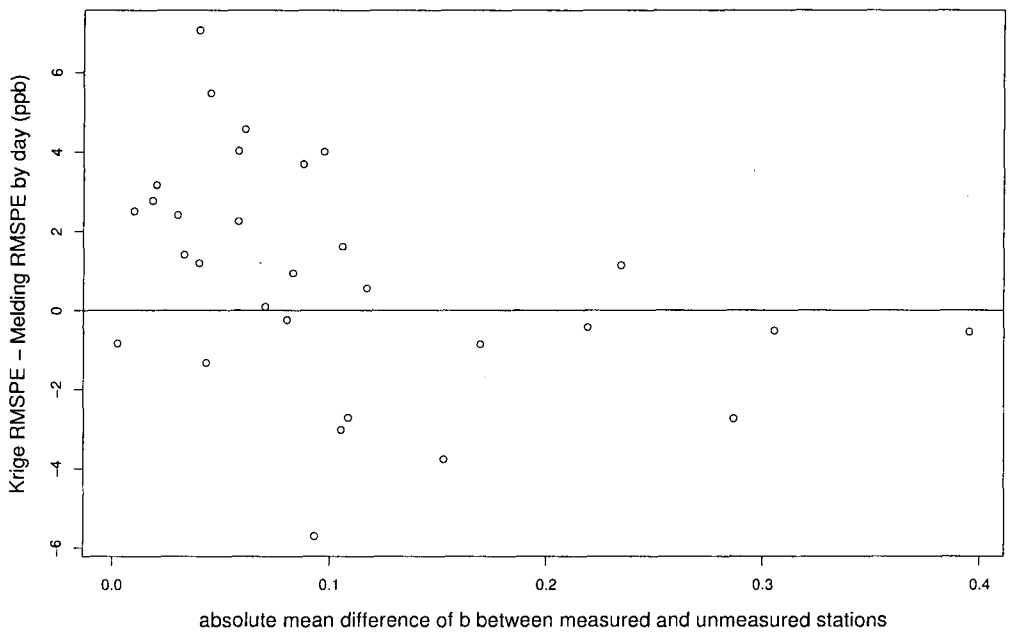

Figure 7.9: Daily RMSPE difference between Kriging with only measurement and the BM model versus the absolute mean difference of $\hat{b}$ (defined by formula (7.4)) between the measured and unmeasured stations. Points above the plotted horizontal line represented victory for the BM model . Notice the supremacy of the BM model when the absolute mean difference of $\hat{b}$ over space is small.

Table 7.2: RMSPE (root mean square prediction error) for predicting daily average of ozone levels at 30 stations. Column 1: days. Column 2: RMSPE for BM model. Column 3: RMSPE for Kriging with measurements. Column 4: RMSPE for Kriging with model outputs. The number with * is the smallest number in each row. The unit of the numbers from column 2 to 4 in this table is ppbv.

day	BM	Kriging 1	Kriging 2
1	14.50	13.98	13.91*
2	8.99	8.58*	12.86
3	11.43	8.71*	11.84
4	13.53	12.19*	15.51
5	13.65	10.63*	15.59
6	14.43	11.69*	16.54
7	13.59	13.04*	13.54
8	11.73*	11.82	12.89
9	17.86	14.11	15.99
10	15.14	9.43	12.76
11	11.59	15.60	11.32
12	15.30*	16.24	15.30
13	14.14*	19.61	15.16
14	15.61*	22.67	17.99
15	18.55	21.06	18.45*
16	17.95	19.36	17.60*
17	11.37*	14.14	11.70
18	13.67	16.84	11.65*
19	6.98 *	11.54	10.51
20	12.84*	14.03	15.45
21	17.44	16.61*	17.33
22	9.48*	11.88	12.54
23	10.04*	11.65	8.65
24	9.60*	9.37	11.99
25	11.76*	12.90	21.11
26	17.55*	18.11	19.03
27	12.26*	15.94	14.95
28	14.31*	18.33	18.07
29	13.49	12.63*	19.10
30	12.41*	14.66	14.26
mean	13.37*	14.24	14.79

week 2 and week 4. On average over all 4 weeks, the BM model has smallest RMSPE.

- As expected, the RMSPE for all the predictors are smaller than RMSPE in the analysis of daily average data.
- the BM's 90% predictive interval has a reasonably good coverage probability of 91.67%.
- The RMSPE for BM in Tables 7.1, 7.2 and 7.3 tells us that BM's predictor improves as the averaging time scale increases. This agrees with the findings in Kasibhatla and Chameides (2000); Hogrefe et al. (2001a), that is, the model outputs represents the measurements better on larger time scales.

7.2 Summary and Conclusions

Previous sections present the details of the BM model and the MCMC algorithm used to fit the model. We have also conducted a comprehensive simulation and applied BM to the ozone data in different time scales.

We see that the BM model has its strengths and weakness. Its strengths are the following.

- The novel method connects the measurement process with the model outputs process by assuming the existence of an underlying process.

Table 7.3: RMSPE (root mean square prediction error) for predicting weekly average of ozone levels at 30 stations. Column 1: days. Column 2: RMSPE for BM model. Column 3: RMSPE for Kriging with measurements. Column 4: RMSPE for Kriging with model outputs. The number with * is the smallest number in each row. The unit of the numbers from column 2 to 4 in this table is ppbv.

week	BM	Kriging 1	Kriging 2
1	9.43	7.98*	13.77
2	9.72*	11.22	10.51
3	10.32	10.21	9.83*
4	8.08*	11.76	12.71
mean	9.38*	10.29	11.71

This approach addresses the difference-in-support problem in a fundamental way.

- the BM model can do a variety of things such as predict unmeasured responses, assessing deterministic model outputs, detecting spatial trends and estimating spatial correlation.

- The model better estimates prediction uncertainty than the classical approach, Kriging. In our data analysis, the coverage probability of the BM's predictive interval is reasonably close to the nominal level.

- Its Bayesian framework makes the BM model relatively easy to extend to incorporate other things like ensembles, the reversible jump MCMC and non-stationary spatial correlation.

However, the BM model also has some weakness as following.

- The computational price is high. By sampling points within grid cells, the dimension of the spatial correlation matrix can be very big even with a modest number of grid cells. Inverting the spatial covariance matrix three times in each MCMC iteration takes a lot of computation time.

- the BM model does not yet cover space-time processes and hence it cannot "borrow strength" over time. Ozone data are recorded hourly, which obviously has strong periodicity and strong auto-correlation. The main challenge to create a "space-time" BM model lies in the

computation burden. With temporal correlation, the space-time correlation matrix will be huge and hence more likely to be ill-conditioned. Inverting such a matrix will be both difficult and computationally expensive. However, the dynamic BEM model we present later in Chpater 10 provides a simple extension to a 'space-time" BM model.

- the BM model's normality assumption poses a problem. For the measurements, a transformation maybe used to validate that assumption. However, with the model outputs data, non-linear transformations cannot be used since the model outputs is represented by such an integral of the true process.

- The locations of sampling points within grid cells change because of the random sampling scheme. Because the mean of the underlying process depends on its geographical coordinates in general (like universal Kriging), the prediction result have some variation every time the BM model is used for the same data.

Finally, we offer the following observations.

- Where some of the stations are within grid cells, it would seem better to use them instead of sampling points within the grid cells to reduce the dimension of the spatial correlation matrix.

- The meaning of the concept of a "true underlying process" seems unclear since it is merely a conceptual construct rather than a physically meaningful process. Thus its may will be criticized.

- We assume the underlying process mean to be a polynomial function of the geographical coordinates, which may not be enough if other variable such as temperature also affects the ozone level. We can side-step the assumption of a true underlying process by regressing the measurements on the model outputs directly.

7.3 Spatial-temporal Model Prediction and Forecasting

In previous sections of this chapter, we have used the BM to predict hourly ozone concentrations and found that its predictions are only marginally better than Kriging. Moreover, the hourly data has strong auto-regression structure which is totally ignored in BM. This section presents a new analysis of the hourly data by using the Bayesian spatial-temporal model proposed in Chapter 5. We still use the hourly measurements and model outputs from the same 78 stations and grid cells as in the previous section. By our selection, each grid cell has one and only one station inside. We use the first 240 hours of measurements and model outputs at 15 stations as training data. Then we forecast the ozone concentration levels in the next 240 hours for the 15 stations and also predict the first 240 hours' ozone concentration levels at the remaining 63 stations. To forecast and spatially predict the ozone concentration levels, we use both the ad-hoc approach and the Bayesian spatial-temporal model proposed in Section 5.2 and Section 5.3. The comparison

between the ad-hoc approach and the Bayesian spatial-temporal model will be summarized in this section.

As a preliminary step, we use the two-step regression procedure. Figure 7.10 shows the auto-correlation function (ACF) and partial auto-correlation function (PACF) of the residuals ϵ_t at only one station. It is clear that the residuals are temporally independent. The ACF and PACF plots of the residuals at other stations show similar results. So the AR(1) temporal structure seems appropriate to model the process $N_{\boldsymbol{s},t}$ in model 5.5.

To fit the Bayesian spatial-temporal model (5.6), we use 500 iterations of Gibbs sampling algorithm and the first 50 iterations for the "burn-in" period. Figure 7.11 shows the time series plots of the posterior MCMC samples for some parameters as examples. From that plot, we can see the Markov Chains converge after the "burn-in" period. We use the root mean square forecast error (RMSFE) and root mean square prediction error (RMSPE) to evaluate accuracy of the forecasts or predictions, the smaller RMSFE or RMSPE, the better the forecast or prediction. We define RMSPE in (7.3). At each station location $\boldsymbol{s}$, RMSFE is defined as

$$\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(O_{\boldsymbol{s},t}-\hat{O}_{\boldsymbol{s},t}\right)^2}, \tag{7.5}$$

$O_{\boldsymbol{s},t}$ being the real measurement at time t and $\hat{O}_{\boldsymbol{s},t}$, the forecast or prediction.

Table 7.4 presents estimation results for some of the parameters in model (5.5). Most of the estimates differ significantly from 0 except σ_a^2, the spatial

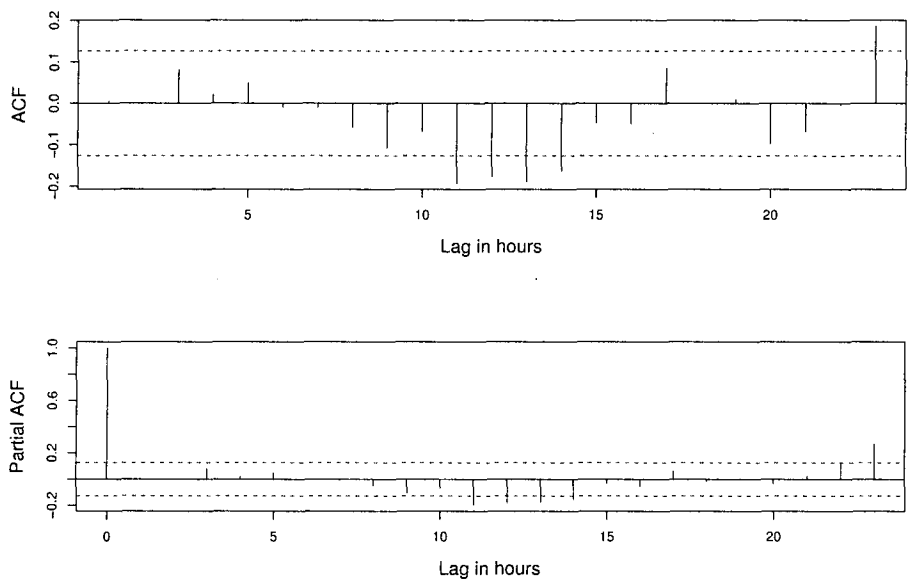

Figure 7.10: Plots of partial ACF and ACF for the estimated residual $\hat{\epsilon}_t$ at one station.

variance of the process $\boldsymbol{a}$. Table 7.4 shows the posterior standard deviations of λ_ϵ to be much smaller than those of λ_a and λ_c. As stated in Section 4.1, the range parameter of the spatial process is usually difficult to estimate. That is why the posterior standard deviations are large for λ_a and λ_c. However, for λ_ϵ, we have independent replicates of the residual process $\{\boldsymbol{\epsilon}_s\}$ over 239 hours in this data analysis, and thereby see a reduction in the posterior standard deviations of λ_ϵ. The RMSFE of the forecasts by three approaches, Bayesian spatial-temporal model (5.5), the ad-hoc approach in Section (5.2) and the model outputs are in Table 7.5. As an example, Figure 7.12 shows the plot of measurements versus the forecast by the Bayesian spatial-temporal model for one station. We also tried to use AR(2) for $N_{\boldsymbol{s},t}$ in model (5.5). The estimation of the second auto-regression coefficient is as small as 0.025 and the standard error of that estimation is 0.013. So the second auto-regression coefficient is not significant. This shows that AR(1) is enough for $N_{\boldsymbol{s},t}$, which in agreement with our preliminary analysis. In additional, the predictions and forecasts become a little worse if we change from AR(1) to AR(2). So for this data analysis, AR(1) is more appropriate than AR(2) for $N_{\boldsymbol{s},t}$.

The spatial predictor predicts the ozone concentration levels during the first 240 hours at the 63 sites that have only model outputs but no historic measurements. The four predictions include the Bayesian spatial-temporal model, the ad-hoc approach, Kriging, and the model outputs. By Kriging, we mean using the Kriging approach to interpolate the measurements at the 63 sites repeatedly for every hour. So the Kriging approach does not use

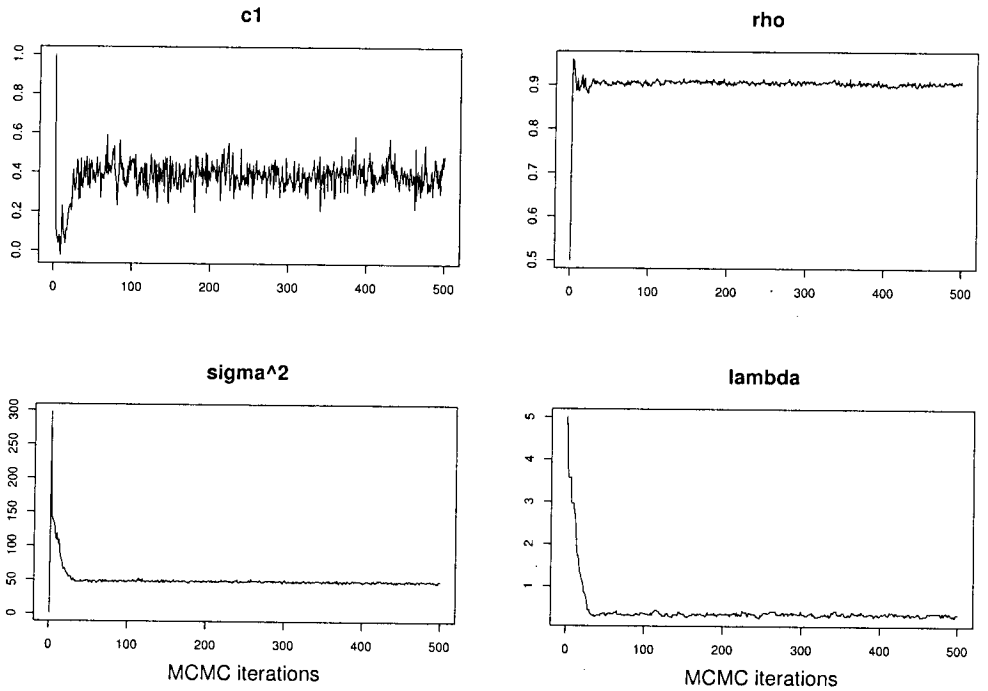

Figure 7.11: Plots of the parameter values in the Gibbs sampling iterations. Upper left panel: a_1; upper right panel: rho; lower left panel: σ_ϵ^2; lower right panel: λ_ϵ.

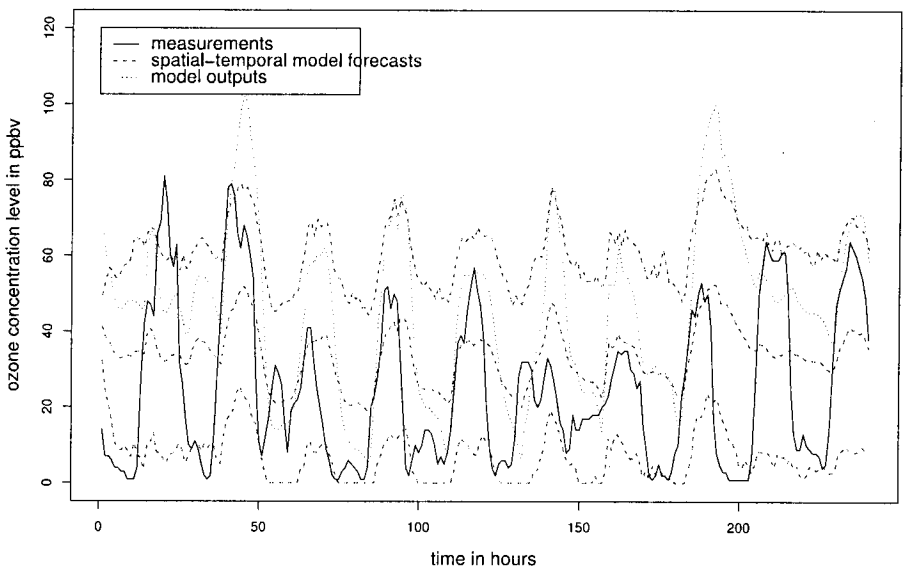

Figure 7.12: Plot of measurement and the forecast and with 90% credible interval using the Bayesian spatial-temporal model at one of the 15 stations.

the model outputs. Tables 7.6 and 7.7 present the RMSPE of these four approaches. As an example, Figure 7.13 shows the plot of measurements versus the spatial predictions using the Bayesian spatial-temporal model at one site.

The Bayesian spatial-temporal model yields credible predictive intervals for the forecasts and predictions through the quantiles of the MCMC samples of the forecasts and predictions. The coverage probability of the credible interval is defined as the proportion of the true measurements that are forecast or predicted, falling in the respective credible intervals. The coverage probabilities of the 90% credible intervals for the forecasts and predictions are 90.00% and 94.69% respectively.

Based on the above forecast and spatial prediction results, we have the following conclusions.

- The two-step linear regression procedure can only forecast the measurements. The Kriging approach can only spatially predict the measurements. Both the ad-hoc approach and the Bayesian spatial-temporal model can both forecast and spatially predict the measurements at the same time. From Figure 7.12 and Figure 7.13, we can see that the forecast and prediction from the Bayesian spatial-temporal model track the real measurements much better than the model outputs.

- Tables 7.5, 7.6 and 7.7 show that in averaging over the stations, the Bayesian spatial-temporal model gives the smallest RMSFE and RM-

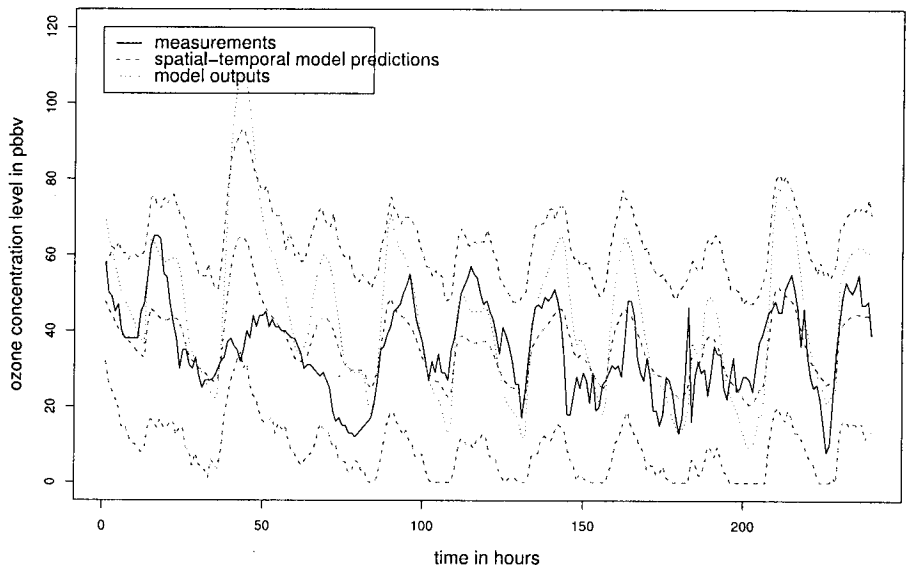

Figure 7.13: Plot of measurement and the spatial prediction with 90% credible interval from the Bayesian spatial-temporal model at one of the 63 ungauged sites.

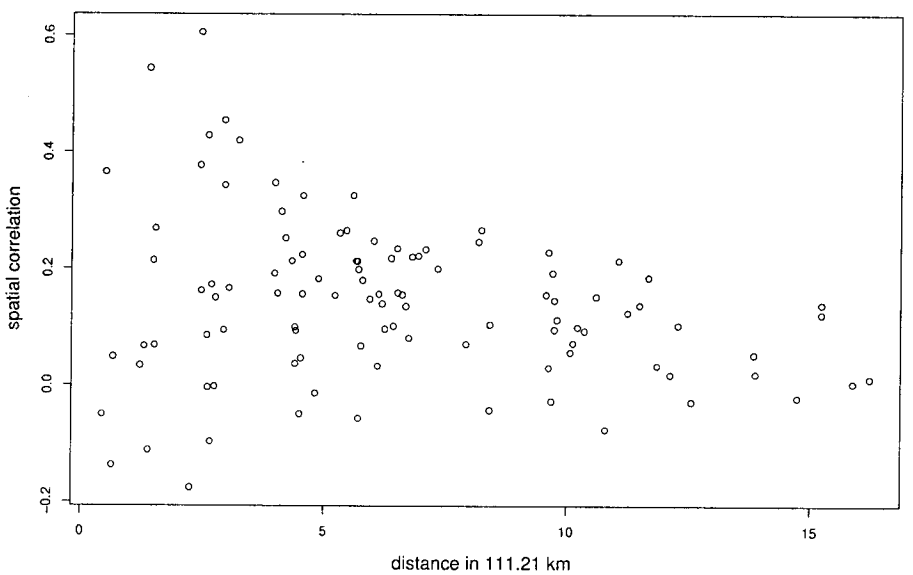

Figure 7.14: Plot of distance versus correlation for the normalized residuals from model 5.5.

SPE among all the competitive approaches for forecasting and spatially prediction. For forecasting, the mean RMSFE of the Bayesian spatial-temporal model is about 16% smaller than the forecasts from the model outputs. The RMSFE of the ad-hoc approach is slightly bigger than the Bayesian spatial-temporal model. In prediction, not surprisingly, Kriging has the biggest RMSPE because we have interpolated 63 sites using measurements from only 15 stations. The RMSPE of the Bayesian spatial-temporal model is about 12% smaller than the ad-hoc approach and the model output. So, using model (5.5) to calibrate the model outputs seems worthwhile for achieving better forecasts and predictions. The coverage probabilities of the intervals for the Bayesian spatial-temporal model are fairly well calibrated, meaning the uncertainties in the forecasts and predictions are characterized for quite well.

- Although in averaging over the sites, the Bayesian spatial-temporal model achieves the smallest RMSFE and RMSPE, it is also marginally better than the ad-hoc approach in forecasting albeit at significant cost in computational time and parsimony judging from the number of parameters. Table 7.5 shows that the Bayesian spatial-temporal model has smaller RMSFE than ad-hoc approach at 9 out 15 stations. Table 7.6 and 7.7 show that the Bayesian spatial-temporal model has smaller RMSFE than ad-hoc approach at 46 out 63 stations. So, recognizing the cost stated above, the major benefit of the Bayesian spatial-

temporal model lies in the higher quality of spatial predictions and in its well calibrated credible predictive intervals.

- Table 7.4 shows the posterior mean of the auto-regressive parameter ρ to be as big as 0.91, which suggests strong auto-regression in the residuals $\{N_t\}$.

- At time t, the residuals vector over space is $\boldsymbol{\epsilon}_t$ and the normalized residuals vector is $\boldsymbol{\epsilon}_{t'} = \boldsymbol{\Sigma}_{\epsilon}^{-\frac{1}{2}}\boldsymbol{\epsilon}_t$. Figure 7.14 shows the empirical spatial correlation for the normalized residuals from the Bayesian spatial-temporal model 5.5 in relationship to inter-site distance. We can see that except for few pairs of stations, the normalized residuals do not have much spatial correlation. There are only two pairs of stations whose normalized residuals still have spatial correlation higher than 0.5. The reason could be the similar temporal pattern of the normalized residuals which are not captured by the spatial-temporal model.

7.4 Multivariate Spatial-temporal Model Prediction and Forecast

This section presents the results of applying the multivariate spatial-temporal model (6.1) and (6.2) to the ozone pollution data described in Section 7.1. We use the measurements and model outputs from the same 78 stations and grid cells as in Section 7.3. We have 119 days of hourly measurements

and model outputs. On each day, we take the 8-hour day time average of measurements and model outputs from 1000 to 1700 and the night time average from the rest hours. On each day, we have the day- and nighttime averages of measurements as bivariate responses. The day- and nighttime averages of model outputs are the covariates $\boldsymbol{M}$ in model (6.1). We use the data for the first 80 days at 15 stations to fit the the multivariate spatial-temporal model. Then we forecast the measurements in the next 39 days at the 15 stations and spatially predict the measurements at the remaining 63 sites for the first 80 days. In model (6.1), we assume the residuals $\{\boldsymbol{N}_{s,t}\}$ at each site form an AR(1) process which has a non-zero mean with covariates $\{\boldsymbol{Z}_{s,t}\}$. In the application of univariate model (5.5), the covariates $\{\boldsymbol{Z}_{s,t}\}$ are indicator functions to include the 12-hour periodicity of the measurements. However, the day- and nighttime average measurements are on a daily scale and they do not have such strong periodicity as the hourly measurements. We have used indicator functions as in Section 5.1 to include the 7-day periodicity of the measurements. However, the forecasts or predictions are not improved. So we just assume $\boldsymbol{\gamma}_s = \mathbf{0}$ in model (6.1) when we do the data analysis. We still include the covariate matrix $\boldsymbol{Z}$ in the multivariate model (6.1) for the completeness and flexibility of the model. We evaluate the forecast and prediction results by RMSPE and RMSFE defined by (7.3) and 7.5. Table 7.8 presents the estimation results for the parameters $\boldsymbol{\mu}_\beta$, $\boldsymbol{\rho}$ and $\boldsymbol{\Sigma}$. An alternative to using the multivariate model (6.1 is to apply the univariate model (5.5) to the day- and nighttime average data separately. To illustrate

the benefit of the multivariate model, we compare the RMSFE and RMSPE of these two approaches. Table 7.9 presents the comparison results.

Table 7.8 and Table 7.9 lead us to the following conclusions.

- The posterior mean of $\mu_{\beta,1,2}$ is significantly different from 0. This shows the model outputs are helpful in predicting and forecasting daytime measurements.

- The posterior means of $\mu_{\beta,2,2}$ and $\mu_{\beta,2,3}$ are not significantly different from 0 because the ozone levels at night are much lower than the daytime. This shows the nighttime measurements are much more difficult to predict or forecast from the model outputs.

- The posterior mean of $\Sigma_{1,1}$ is much larger than $\Sigma_{2,2}$. In other words, the variation of daytime measurements are much larger than the nighttime measurements. This is not surprising to us because the ozone concentration levels are much higher at the daytime.

- The coefficients of the auto-regressive matrix $\boldsymbol{\rho}$ are significantly different from 0. However, the auto-regression is not as strong as in the univariate case. The estimation of the auto-regressive parameter in univariate case is 0.91 while in multivariate case the parameters are no larger than 0.45. The reason is that the measurements are on an hourly scale in the univariate analysis but are daily averages in multivariate analysis.

- Table 7.9 shows that the multivariate model forecasts and predictions have smaller RMSPE and RMSFE than the univariate models for both day- and nighttime measurements. So it is worth using multivariate spatial-temporal model (6.1) to achieve better forecasts and predictions of the day- and nighttime measurements for daily measurement aggregates.

- The coverage probabilities for 90% prediction and forecast credible intervals are 84.73% and 83.93% respectively. This shows the uncertainties of the predictions and forecasts are taken into account reasonably well by the model.

Table 7.4: Posterior means and standard deviations of parameters $a_1, \cdots, a_5$, $c_1, \cdots, c_5$, σ_ϵ^2, λ_ϵ, λ_a, σ_a^2, σ_c^2, λ_a and ρ. Since this table is for the estimates of parameters, there is no unit associated to the numbers in this table. Each of these parameters is a scalar.

	mean	sd		mean	sd
a_1	16.57	2.81	a_2	16.67	2.83
c_1	0.38	0.07	c_2	0.38	0.06
σ_ϵ^2	48.09	1.15	a_3	20.62	3.89
λ_ϵ	0.35	0.03	c_3	0.35	0.06
λ_a	4.70	2.41	a_4	19.73	3.40
σ_a^2	33.00	27.4	c_4	0.37	0.07
σ_c^2	0.33	0.13	a_5	17.48	2.92
ρ	0.91	0.004	c_5	0.62	0.08

Table 7.5: The first column gives the indeces of 15 stations where the first 240 hours' measurements are available. The second, third and forth columns are the RMSFE (square root of the mean square forecast error) of the next 240 hours' measurements fore-casted by the Bayesian spatial-temporal model, ad-hoc approach and model outputs. The number followed by $*$ indicates the approach that has the smallest RMSFE in that row. The unit of the numbers from column 2 to 4 in this table is ppbv.

station	spatial-temporal	ad-hoc	model outputs
1	21.51*	21.51*	26.30
2	19.70*	20.99	31.13
3	13.73	12.02*	16.29
4	18.16	16.69*	20.06
5	15.61*	17.21	14.36
6	14.54	10.51*	16.51
7	19.33*	22.12	26.90
8	19.91	19.30*	18.98
9	14.08	13.67*	16.74
10	11.28*	12.57	11.58
11	12.04*	12.70	16.64
12	15.37	15.09*	15.89
13	10.95*	12.20	15.04
14	12.57*	14.32	14.66
15	14.71*	15.63	17.51
mean	15.57*	15.77	18.57

Table 7.6: RMSPE at ungauged sites 1-35. Column 2: Bayesian spatial-temporal model; column 3: ad-hoc approach; column 4: model outputs; column 5: Kriging with measurements. The number followed by $*$ indicates the approach that has the smallest RMSPE in each row. The unit of the numbers from column 2 to 5 in this table is ppbv.

site	spatial-temporal	ad-hoc	model	Kriging
1	16.16	13.08*	16.32	16.51
2	11.68*	12.07	13.48	17.05
3	13.19 *	13.91	13.38	24.15
4	16.17*	17.08	22.65	19.17
5	14.63*	29.98	19.59	43.40
6	14.17	18.15	11.85*	40.89
7	12.59 *	12.98	14.55	22.80
8	12.90*	13.62	17.35	23.10
9	16.91*	17.28	21.85	27.81
10	14.74	16.26	13.31*	31.90
11	15.31	16.07	11.49*	35.59
12	15.82	17.94	15.02*	29.32
13	15.33	20.35	11.77*	36.71
14	12.75*	15.37	16.46	20.84
15	12.62	11.49*	15.05	16.70
16	14.71	12.91*	14.24	17.11
17	19.49	18.77	18.35*	19.14
18	11.35*	15.76	16.15	23.07
19	16.44	14.29*	14.76	18.33
20	15.83	14.15*	15.38	17.70
21	12.85	13.06	12.65*	18.79
22	11.66	10.26 *	11.21	14.83
23	15.56	16.25	16.11	10.38*
24	9.44*	10.75	14.17	19.64
25	15.08*	15.93	15.47	24.78
26	14.08	13.40	13.33*	25.74
27	11.41*	19.10	15.94	27.33
28	12.42*	12.94	14.97	26.94
29	16.00 *	19.17	18.02	32.11
30	12.45*	18.70	13.94	32.67
31	19.24	17.84*	21.82	25.94
32	14.22*	21.61	16.28	34.19
33	14.04*	22.79	16.47	32.95
34	14.91*	15.29	15.39	27.05
35	12.40	14.76	10.86*	24.43

Table 7.7: RMSPE at ungauged sites 36-63. Column 2: Bayesian spatial-temporal model; column 3: ad-hoc approach; column 4: model outputs; column 5: Kriging with measurements. The number followed by $*$ indicates the approach that has the smallest RMSPE in each row. The unit of the numbers from column 2 to 5 in this table is ppbv.

site	spatial-temporal	ad-hoc	model	Kriging
36	13.65*	22.56	16.01	35.43
37	14.29*	23.24	18.30	33.09
38	12.71	12.52*	13.91	26.54
39	18.22*	21.30	24.27	20.20
40	13.75*	14.86	18.99	16.91
41	13.80*	14.49	20.00	16.57
42	13.93*	14.72	18.46	18.77
43	19.59	14.66*	19.99	26.53
44	19.39	14.28*	17.69	23.84
45	15.12	15.07	17.79	14.22*
46	13.54*	14.35	13.75	31.88
47	13.40*	13.80	15.21	28.34
48	16.18	15.54*	18.91	19.62
49	15.02*	20.88	19.65	25.71
50	12.39*	15.35	13.42	13.38
51	11.24*	11.48	12.86	14.31
52	11.78*	12.60	16.20	13.79
53	17.99	18.07	19.14	16.47*
54	13.04*	12.77	18.63	13.77
55	15.08*	15.47	17.42	15.31
56	11.98*	17.61	18.32	21.02
57	13.98*	19.41	20.09	25.79
58	14.70*	19.02	17.01	25.05
59	11.01*	17.40	18.97	23.48
60	10.00*	13.34	18.36	18.03
61	14.23*	14.32	16.15	19.83
62	22.87	18.49	24.94	14.62*
63	17.97	15.48*	15.92	19.99
mean	14.43*	16.13	16.50	23.36

Table 7.8: Posterior mean and standard deviations of the parameters μ_β, ρ and Σ. $\mu_{\beta,1,1}$ and $\mu_{\beta,2,1}$ are the means of intercepts; $\mu_{\beta,1,2}$ and $\mu_{\beta,1,3}$, the mean of coefficients of day- and nighttime model outputs of the daytime measurements; $\mu_{\beta,2,2}$ and $\mu_{\beta,2,3}$, the mean of coefficients of day- and nighttime model outputs on the nighttime measurements. Since this table is for the estimates of parameters, there is no unit associated to the numbers in this table. Each of these parameters is a scalar.

	mean	sd		mean	sd
$\mu_{\beta,1,1}$	17.82	3.72	$\rho_{1,1}$	0.45	0.01
$\mu_{\beta,1,2}$	0.41	0.09	$\rho_{1,2}$	0.22	0.01
$\mu_{\beta,1,3}$	0.20	0.09	$\rho_{2,1}$	0.17	0.02
$\mu_{\beta,2,1}$	17.54	1.27	$\rho_{2,2}$	0.25	0.01
$\mu_{\beta,2,2}$	0.05	0.10	$\Sigma_{1,1}$	131.89	-
$\mu_{\beta,2,3}$	0.11	0.10	$\Sigma_{1,2}$	10.33	-
$\Sigma_{2,1}$	10.33	-	$\Sigma_{2,2}$	50.37	-

Table 7.9: Average of RMSPE and RMSFE for Bayesian multivariate spatial-temporal model (6.1) (columns 2 and 4) and univariate spatial-temporal model (5.5) (columns 3 and 5). The average of RMSPE is over 63 sites and the average of RMSFE is over 15 stations. The unit of the numbers in this table is ppbv.

	RMSPE		RMSFE	
	multivariate	univariate	multivariate	univariate
daytime	16.78	18.10	14.44	15.91
nighttime	12.51	14.19	9.69	9.71

Chapter 8

Calibration of Deterministic models

Chapter 1 describes the ozone's policy relevant background (PRB) level and why it cannot be measured. That difficulty forces use of deterministic model output to infer the PRB level. However, that output needs to be calibrated to ensure the deterministic model output represents the PRB level. Calibration needs to be based on the measurements. This chapter explains how we can use the BM model (2.9) and Bayesian spatial-temporal model (5.5) to calibrate deterministic models and it presents an example involving model outputs of ozone concentrations.

As in Chapter 7, we focus on the 8-hour daytime average measurements and model outputs for 78 grid cells and stations during the 30 days from July 1 to July 30, 1995. The calibrated model output at grid cell B by using BM model is

$$\left(\tilde{Z}(B) - \frac{1}{|B|}\int_B a(\boldsymbol{s})d\boldsymbol{s}\right)/b. \tag{8.1}$$

The multiplicative calibration parameter b is assumed constant across space

and $a(\boldsymbol{s}) = a_0 + a_1 s_1 + a_2 s_2$ while s_1 and s_2 are the longitude and latitude of location $\boldsymbol{s}$. Fuentes and Raftery (2005) plug-in the posterior mean of $a(\boldsymbol{s})$ and b into (8.1) to obtain the calibrated model output.

Departing the plug-in approach used by Fuentes and Raftery (2005), we calibrate the model output using the BM model directly. In each iteration of the Gibbs sampling, we use formula (8.1) to calculate the calibrated model output, so that we get the empirical distribution needed for calibration. The calibrated model output by using the Bayesian spatial-temporal model is

$$a + c\tilde{Z}(B). \tag{8.2}$$

To illustrate the benefit of calibrating the model output, we compare in Table 8.1 the RMSPE for the calibrated and uncalibrated model outputs for the 30 days on which our study focuses. As the daily data analysis in Section 7.1. The data available is the measurements at 48 stations and model outputs at all the 78 grid cells. The calibrated model output at the remaining 30 stations are used to predict the measurements there. As an example, Figure 8.1 shows the plot of the measurements versus the uncalibrated and calibrated model output for day 1. Both the BM and Bayesian spatial-temporal models calibrated model outputs are included in that plot. Figure 8.2 shows the scatter plots of BM model calibrated model outputs versus uncalibrated model outputs and Figure 8.3 shows scatter plots of Bayesian spatial-temporal model calibrated model outputs versus uncalibrated model

outputs. From the calibration results for the model output, we have the following conclusions.

- Figure 8.2 and 8.3 reveal marked differences between calibrated and uncalibrated model outputs because many points deviate from the solid line with intercept 0 and slope 1.

- Table 8.1 shows that on average the Bayesian spatial-temporal model calibrated model outputs have the smallest RMSPE than the uncalibrated ones and BM calibrated model outputs also have smaller RMSPE than uncalibrated ones. After calibration, the mean RMSPE has been reduced by 16.07% for BM model, 21.48% for Bayesian spatial-temporal model. Out of all the 30 days, both BM and Bayesian spatial-temporal model calibrated model outputs have smaller RMSPEs than uncalibrated ones for 25 days. So it is beneficial to calibrate the model outputs by using either BM or Bayesian spatial-temporal model.

- Figure 8.1 shows that both BM and Bayesian spatial-temporal calibrated model outputs are closer to the measurements than the uncalibrated ones at most stations.

- Figure 8.4 shows the histograms of posterior means of calibration parameters a_0, a_1, a_2 and b in the 30 days. The posterior means of b range from 0.9 to 2.0 in most of the days.

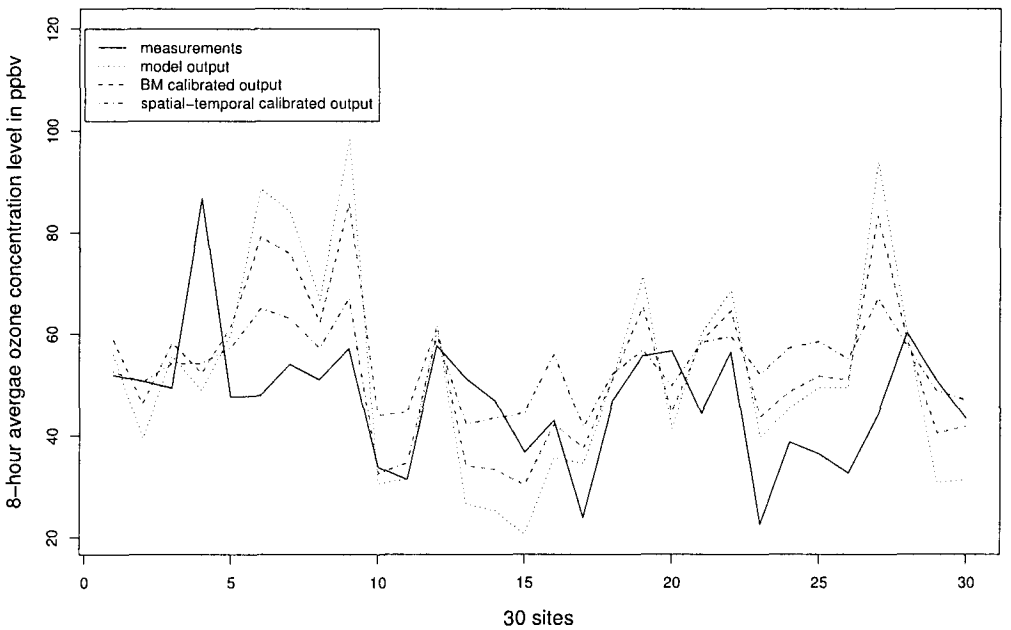

Figure 8.1: Measurements versus the uncalibrated and calibrated model output of 30 grid cells on day 2. The solid line shows the measurements inside the grid cells. The dotted and dashed lines are the uncalibrated and calibrated model outputs respectively.

Table 8.1: RMSPE of uncalibrated and calibrated model outputs for the prediction of measurements at the 30 stations. column 1: day 1-30; column 2: RMSPE of uncalibrated model output; column 3: RMSPE of BM calibrated model output; column 4: RMSPE of Bayesian spatial-temporal model calibrated model output; The number with a * indicates the "winner" in that row. The unit of the numbers from column 2 to 4 in this table is ppbv.

Day	model output	BM	spatial-temporal
1	20.43	16.54	13.39*
2	19.88	10.89*	12.17
3	10.41	7.60*	8.78
4	14.82	13.79	12.00*
5	22.72	16.05	15.94*
6	23.67	15.09	14.02*
7	16.14	13.75	13.12*
8	15.36	14.42	11.30*
9	14.66	14.61*	15.06
10	12.23	11.19*	12.42
11	12.96	10.01	9.78*
12	17.41	16.19	15.00*
13	15.07	18.45	15.25*
14	20.83	15.77*	22.28
15	23.19	20.14*	22.79
16	16.49	17.28	16.22*
17	15.68	11.47	11.09*
18	15.89	15.35	13.83*
19	10.30	10.44	7.29*
20	14.20	11.59*	13.14
21	19.78	26.13	14.37*
22	17.96	11.07	8.98*
23	12.29	11.37	8.29*
24	15.87	10.02	9.53*
25	22.67	12.35*	13.92
26	18.41	14.06 *	14.12
27	16.59	9.99*	10.80
28	19.70	18.75	13.15*
29	13.44	11.33*	15.92
30	15.06	17.41	11.69*
mean	16.80	14.10	13.19*

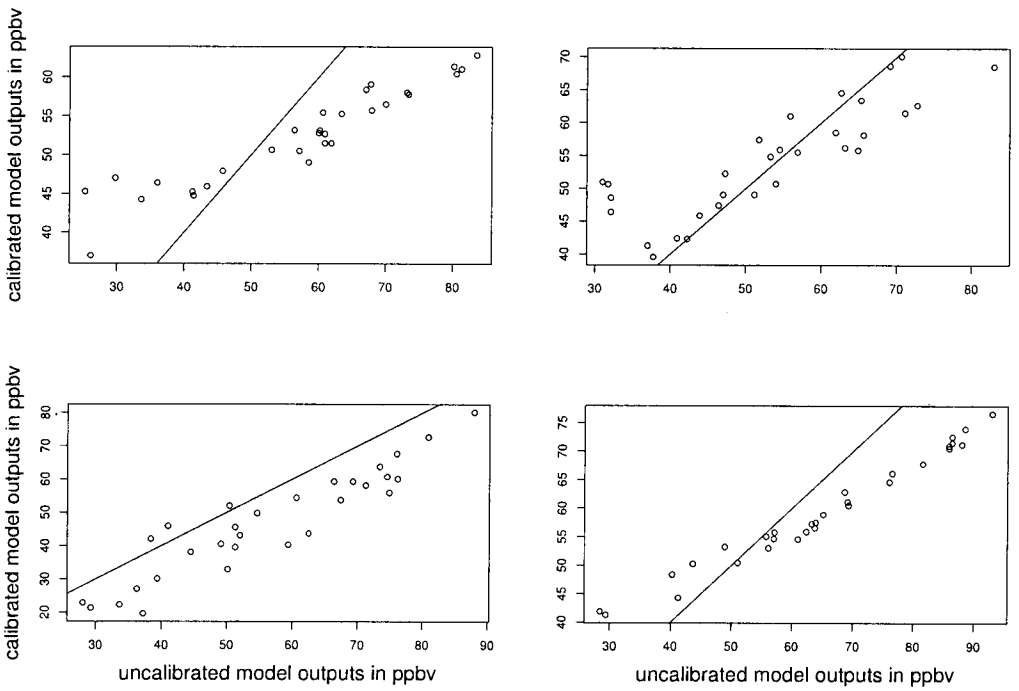

Figure 8.2: Scatter plots of uncalibrated model outputs versus BM calibrated model outputs. The solid lines have intercept 0 and slope 1. The plots are for days 2, 7, 8 and 26 from the upper left to the lower right corner in left-to-right sequence.

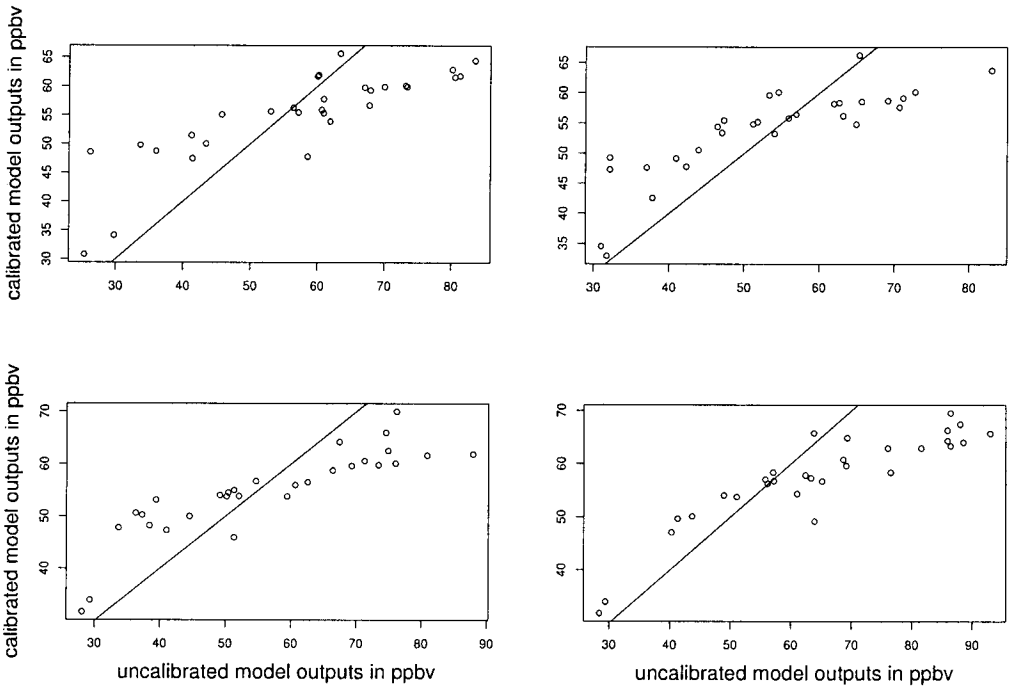

Figure 8.3: Scatter plots of uncalibrated model outputs versus Bayesian spatial-temporal model calibrated model outputs. The solid lines have intercept 0 and slope 1. The plots are for days 2, 7, 8 and 26 from the upper left to the lower right in left-to-right sequence.

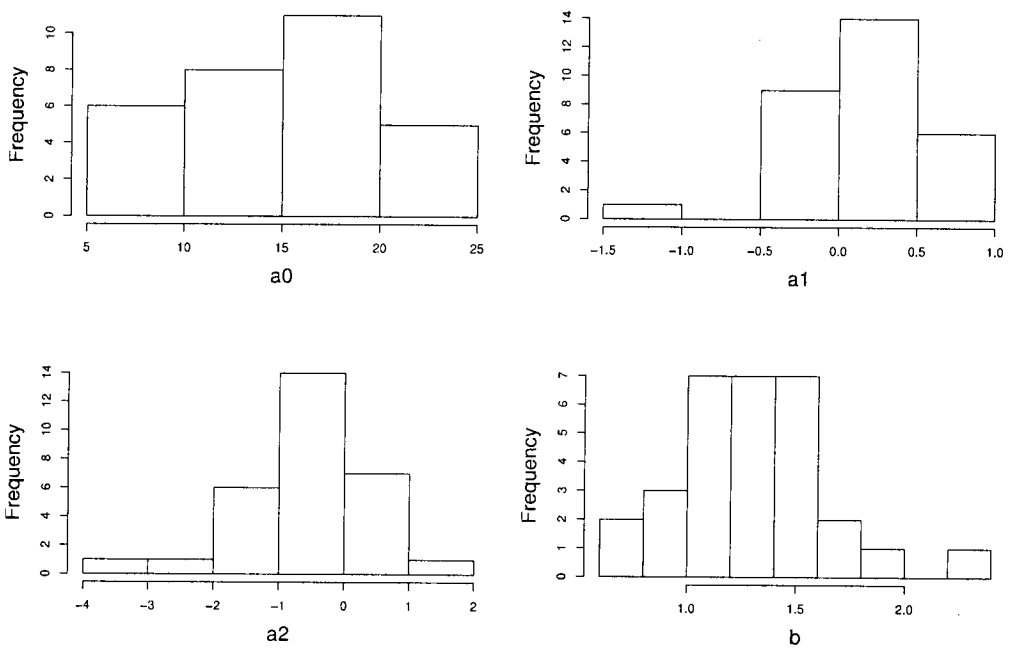

Figure 8.4: Histograms of posterior means of a_0, a_1, a_2 and b during the 30 days on which the current study focuses on.

Chapter 9

Application of the Bayesian Ensemble Melding Model

This chapter presents the results of applying the BEM model to combine measurements and model outputs from an ensemble of deterministic models. Section 3.5 gives a technical introduction to the BEM model. Later in this chapter, we present the additional detailed modifications needed for a particular application we address. Section 9.1 introduces the intended application and Section 9.2 presents the results and conclusions.

9.1 Data Description

Recently, people have paid attention not only to the weather forecast but also to the uncertainties of the forecast. Uncertainty is already a well established concept in statistical community. For example, by using a linear regression model, statisticians not only predict the response based on the available data but they also give a predictive interval for the predicted value. Raftery et al. (2005) propose an idea to probabilistic forecasting and they have already

set up a website (http://probcast.com/) to give such probabilistic forecast for the Pacific Northwest area. In Bayesian models, it is straightforward to use the MCMC samples from the posterior predictive distribution to give a probabilistic forecast. In this section, we give an example of using the BEM model to combine measurements with model outputs from more than one deterministic models to give spatial predictions and their credible intervals.

For this purpose, we downloaded data from the MURI project website (http://www.stat.washington.edu/MURI/). This dataset includes measurements and model output of the sea level temperature in the Pacific Northwest area. The time of the measurements and model outputs are available at 102 non-consecutive hours during January-June 2000. The model outputs are generated by the University of Washington mesoscale short-range ensemble system for the Pacific Northwest, which has been described by Grimit and Mass (2002). This system is a five-member multi-analysis ensemble consisting of different runs of the fifth generation Pennsylvania State University National Center for Atmospheric Research Mesoscale model (MM5). Each run of the MM5 model is made with different initial conditions. The raw deterministic model output are available on grid cells. However, preliminary analysis has been done to interpolate the model outputs to the locations of monitoring stations. So for this dataset, each station has one measurement and five "aligned" model outputs for the temperature. Raftery et al. (2005) modeled this dataset by using Bayesian model averaging of the five model outputs to make the 48-hour ahead forecast of the temperature. Here we

model the same dataset by using ensemble BM for future comparison.

The number of monitoring stations varies from 311 to 650 at different hours. For each hour, we first calculate the Pearson correlation coefficients between measurement and each of the five model outputs over space. Thus, we have 102 correlation coefficients for each of the five initial conditions. Figure 9.1 shows the histograms of these correlation coefficients. Clearly the correlation coefficients are larger than 0.7 at most hours. The strong correlation makes the application of the BEM model a reasonable choice to combine the measurements and model outputs.

9.2 Data Analysis

In our study, the number of stations varies from hour-to-hour and we choose a subset of 150 stations which have measurements and model outputs available at each hour. We use the BEM model to combine measurements from 30 stations and model outputs for all 150 stations to predict the temperature at the remaining 120 stations. We apply the BEM model to the measurements and model output at each hour repeatedly. Unlike the ozone data analyzed in Section 7.1, the model outputs of temperature are available at locations after interpolation and the data file we have does not have the coordinates of the grid cells' vertexes. So in this analysis, we no longer have the concept of grid cells and the conditional distribution (3.10) therefore needs some small modifications. Assume that we have m stations in total and measurements

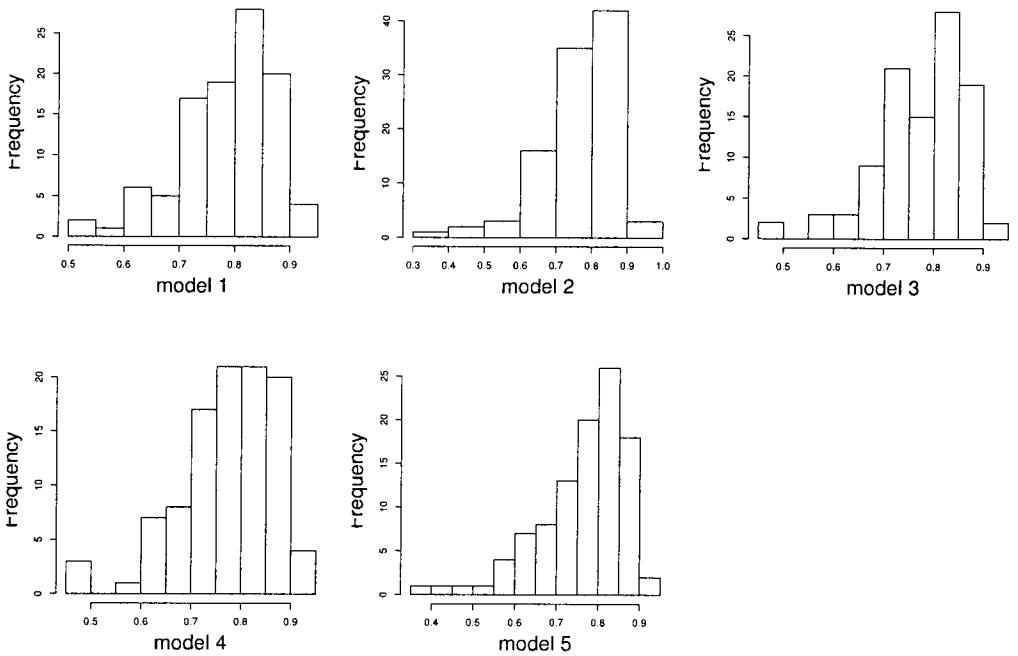

Figure 9.1: Histograms of 102 hours' Pearson's correlation coefficients between measurement and model output over space. Panels 1-5 are the histograms for model output from five different MM5 initial conditions.

are available at n stations, a subset of the m stations. However, the model outputs are available at all m stations. Vector $\boldsymbol{Z}$ is the realization of the true underlying process at the m stations, vectors $\tilde{\boldsymbol{Z}}_i$ $(i = 1, \cdots, 5)$, the realizations of the model output process from i-th deterministic model at the m stations, and vector $\hat{\boldsymbol{Z}}$, the realization of the measurement process at the n stations. Similar to but different from (3.10), the conditional distribution of $(\hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}}_1, \cdots, \tilde{\boldsymbol{Z}}_p, \boldsymbol{Z})$ given $(\boldsymbol{\beta}, \boldsymbol{\theta}, a_1, \cdots, a_p, b_1, \cdots, b_p, \sigma_e^2, \sigma_{\delta,1}^2, \cdots, \sigma_{\delta,p}^2)$ is the following.

$$
\begin{aligned}
& p(\hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}}_1, \cdots, \tilde{\boldsymbol{Z}}_p, \boldsymbol{Z}) | \boldsymbol{\beta}, \boldsymbol{\theta}, a_1, \cdots, a_p, b_1, \cdots, b_p, \sigma_e^2, \sigma_{\delta,1}^2, \cdots, \sigma_{\delta,p}^2) \\
\propto \; & \exp\left\{-\frac{1}{2}\left[\left(\hat{\boldsymbol{Z}} - A_0\boldsymbol{Z}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_1^{-1}\left(\hat{\boldsymbol{Z}} - A_0\boldsymbol{Z}\right)\right.\right. \\
+ \; & \sum_{i=1}^{p}\left(\tilde{\boldsymbol{Z}}_i - a_i - b_i\boldsymbol{Z}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_i^{-1}\left(\tilde{\boldsymbol{Z}}_i - a_i - b_i\boldsymbol{Z}\right) \\
+ \; & \left.\left.(\boldsymbol{Z} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu})\right]\right\} \\
= \; & \exp\left\{-\frac{1}{2}\left[-\boldsymbol{Z}^{\mathrm{T}}\left(A_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\hat{\boldsymbol{Z}} + \sum_{i=1}^{p} b_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\tilde{\boldsymbol{Z}}_i + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)\right.\right. \\
- \; & \left(\hat{\boldsymbol{Z}}^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}A_0 + \sum_{i=1}^{p} b_i\tilde{\boldsymbol{Z}}_i\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{Z} \\
+ \; & \left.\left.\boldsymbol{Z}^{\mathrm{T}}\left(\boldsymbol{\Sigma}_1^{-1} + \sum_{i=1}^{p} b_i^2\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}^{-1}\right)\boldsymbol{Z}\right]\right\}, \qquad (9.1)
\end{aligned}
$$

where $p = 5$ and A_0 is defined as

$$
A_0 = \begin{pmatrix} 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & 0 & \dots & 0 \end{pmatrix}_{n\times m} .
$$

The first n columns of A_0 form an identity matrix and the rest elements of A_0's are 0.

The prior distribution of (a_i, b_i), $i = 1, \cdots, p$, is

$$\begin{pmatrix} a \\ b \end{pmatrix} \sim \mathrm{MVN}(\bar{\boldsymbol{\beta}}_i, \bar{\boldsymbol{F}}).$$

The full conditional distribution of (a_i, b_i) is

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} | \tilde{\boldsymbol{Z}}_i, \boldsymbol{Z}, \boldsymbol{\Sigma}_i \sim \mathrm{MVN}(\bar{\boldsymbol{B}}\bar{\boldsymbol{C}}, \bar{\boldsymbol{B}}),$$

where

$$\bar{\boldsymbol{B}} = \begin{pmatrix} \mathbf{1}^{\mathrm{T}} \\ \boldsymbol{Z}^{\mathrm{T}} \end{pmatrix} (\boldsymbol{\Sigma}_i)^{-1} \begin{pmatrix} \mathbf{1}, & \boldsymbol{Z} \end{pmatrix} + \bar{\boldsymbol{F}}_i^{-1}$$

and

$$\bar{\boldsymbol{C}} = \begin{pmatrix} \mathbf{1}^{\mathrm{T}} \\ \boldsymbol{Z}^{\mathrm{T}} \end{pmatrix} (\boldsymbol{\Sigma}_i)^{-1} \begin{pmatrix} \mathbf{1}, & \tilde{\boldsymbol{Z}}_i \end{pmatrix} + \bar{\boldsymbol{F}}_i^{-1} \bar{\boldsymbol{\beta}}_i,$$

where $\boldsymbol{\Sigma}_i$ is a diagonal matrix with dimensions $m \times m$. The diagonal elements are δ_i^2, the variance of model output error of deterministic model i.

We first assume each additive calibration parameter a is a linear function of the coordinates of location $\boldsymbol{s}$, that is, we have

$$a_s = a_0 + a_1 s_1 + a_2 s_2,$$

s_1 and s_2 being the longitude and latitude of location $\boldsymbol{s}$. However, after fitting the BEM model, we discovered that the coefficients a_1 and a_2 are

not significant. So we assume both the calibration parameters a and b are constants across space. However, the calibration parameters still differ for the different deterministic models. That is, we have five sets of parameters a and b. Figure 9.2 and Figure 9.3 show the histograms of the posterior mean of a and b for different models. Each histogram shows the posterior mean of a and b for 102 hours. To predict the temperature at the 120 stations, we have three competitive purely spatial approaches: BEM, Kriging with measurements and averaging of the five model outputs. Figure 9.4 shows the histogram of the RMSPE of these three approaches over the 102 hours. The above analysis leads us to the following findings.

- The mean RMSPEs of three prediction approaches over 102 hours are 2.98 (BEM), 3.00 (averaging the five model output with equal weights) and 4.10 (Kriging with measurement). The BEM model has the smallest mean RMSPE while the Kriging has the largest. Over the 102 hours, the BEM model achieves smaller RMSPE than Kriging at each of 77 hours. The coverage probability of the BEM's 90% credible intervals for the prediction is 82%, which although reasonable, suggests the BEM slightly underestimate the uncertainties of predictions.

- In Kriging, we use the measurements at only 30 stations to predict 120 stations. So, it is not surprising that Kriging does not predict the measurements very well given the sparsity of the available data. Although the mean RMSPEs between BEM and model averaging outputs is very

close, Figure 9.4 shows that the distribution of BEM RMSPEs is more left skewed than of RMSPEs from simply averaging model outputs. Also BEM can estimate the calibration parameters for each of the five models.

- Figure 9.3 and 9.2 show that the posterior means of b_i are in the interval $[0.9, 1.05]$ and $[0, 20]$ respectively for most hours. The range of the measurements is from 250 to 313 ppbv. So overall, even without calibration, the model outputs represent the measurement very well. This coincides with the strong correlation between measurement and model outputs shown in Figure 9.1.

- Figure 9.5 presents a plot of the RMSPE difference between Kriging and the BEM model versus the mean correlation coefficient between measurement and model output. The mean correlation coefficient is defined as the mean of five correlation coefficients between measurement and five model output. The plot shows that the stronger the correlation between measurement and model outputs, the better melding's predictions compared with Kriging.

To summarize, the BEM model gives the best spatial prediction measured by RMSPE. The BEM model also reflects prediction uncertainty reasonably well. So, with a small number of measurements but a large number of model outputs, we can combine the measurements and model outputs to achieve better spatial prediction. The model outputs help in prediction because of

the strong correlations between measurement and model output.

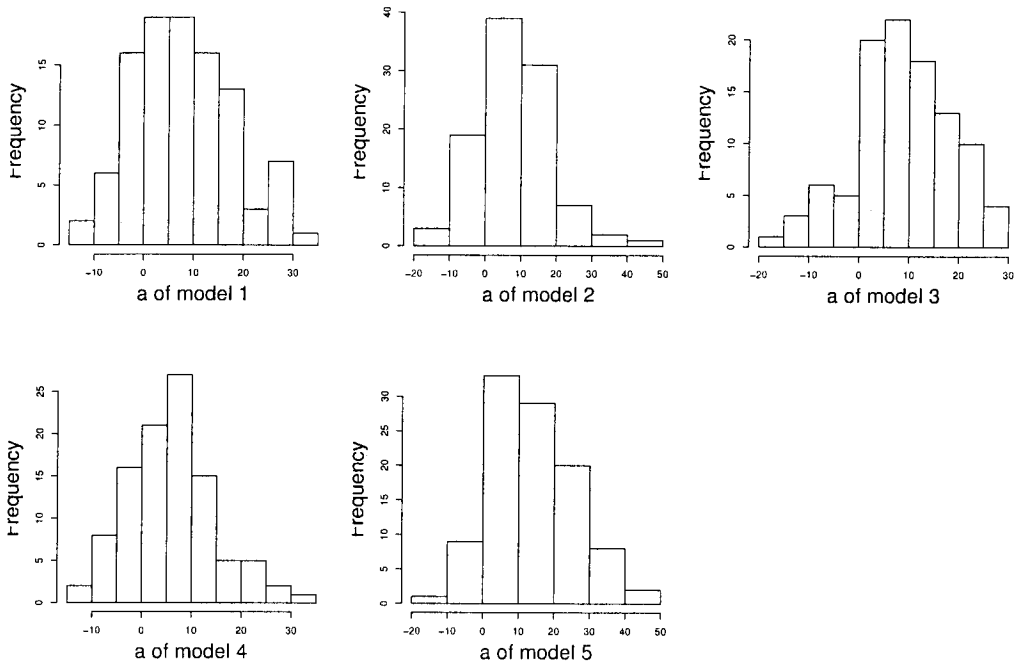

Figure 9.2: Histograms of posterior mean of a. Panels 1-5 are the histograms of a for model output from five different MM5 initial conditions.

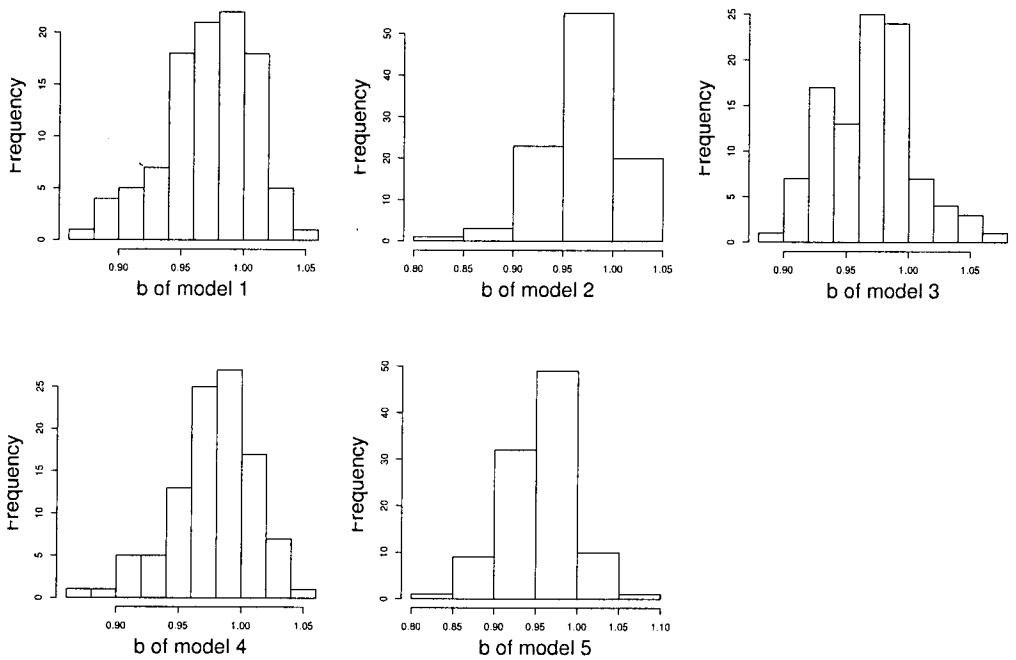

Figure 9.3: Histograms of posterior mean of b. Panels 1-5 are the histograms of b for model output from five different MM5 initial conditions.

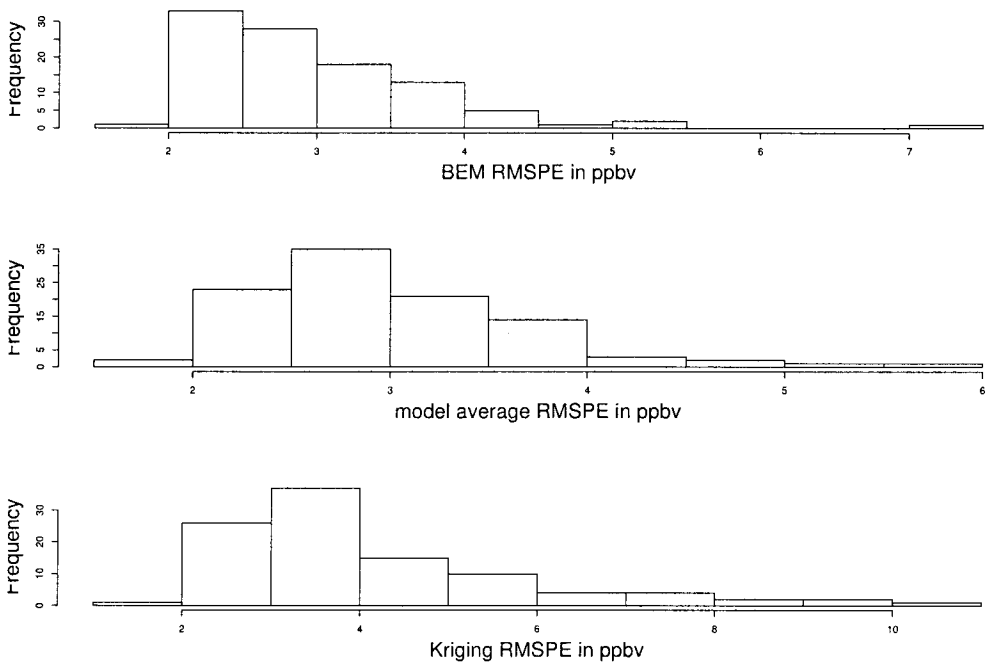

Figure 9.4: RMSPE histograms of predictions by BEM, model averaging with equal weight and Kriging with measurements only.

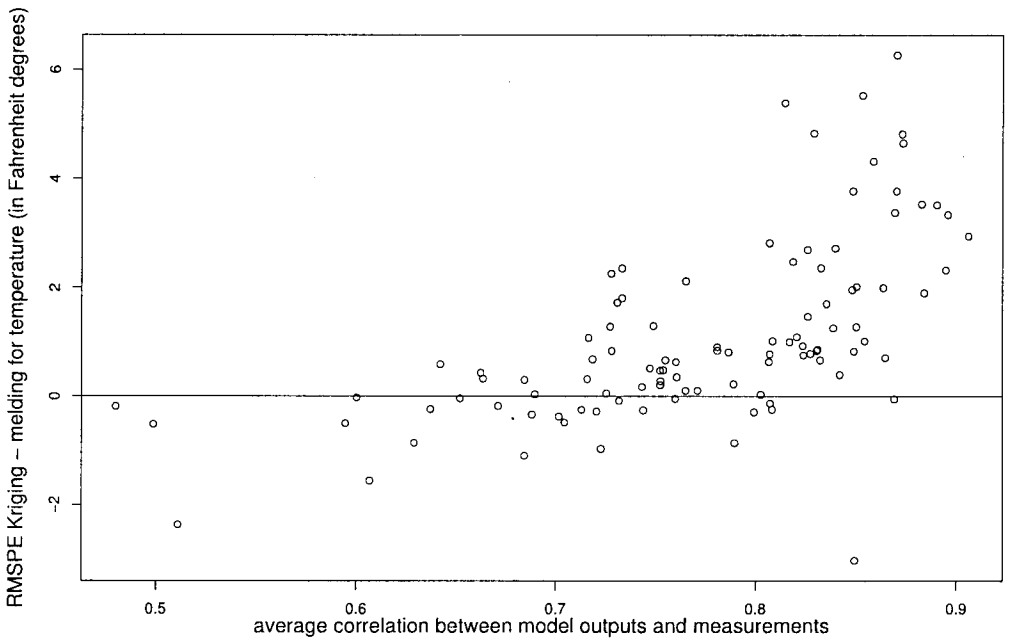

Figure 9.5: The x-axis represents the mean correlation coefficients between measurement and model output, i.e. the mean of the five correlation coefficients between the measurement and five model outputs.

Chapter 10

Dynamic Bayesian Ensemble Melding Model

Both the BM model in Section 2.4 and its extension, the BEM model in Section 3.5 are purely spatial models. So we need to extend them to spatial-temporal models. In some cases, a simple extension is available without the complexities involved in characterizing the temporal structure such as in AR or dynamical linear models. Section 10.1 presents a decision-making problem and its solution. Section 10.2 present a DBEM (dynamic Bayesian ensemble melding) model which can forecast the future with historic data by using the solution of that decision-making problem.

10.1 DeGroot's problem

Bayarri and DeGroot (1989) propose a decision making model to combine opinions from different experts. The idea for this model originally derives from a problem in the book by DeGroot (1970). We present our proof of DeGroot's problem and later show how to construct a dynamic BEM model

by using its conclusion.

Theorem 10.1.1 (DeGroot's theorem) *Each of the K experts has his/her own prior distribution for a certain parameter W, and let π_i be the prior distribution which expert i assigns to W. The executive forms his/her opinion about W from the opinion of the K experts and that his/her prior for W is an weighted average of the π_i. So, the executive's prior for W is*

$$\pi(w) = \sum_{i=1}^{K} \alpha_i \pi_i(w) \quad \sum_{i=1}^{K} \alpha_i = 1, \ \alpha_i \geq 0.$$

Furthermore, the K experts and the executive observe together the value of a random variable X whose conditional distribution when $W = w$ is $f(\cdot|w)$. The posterior distribution of the executive will again be a weighted average of the posterior for the K experts.

Proof The posterior for expert i is

$$\pi_i^*(w|x) = \frac{\pi_i(w) f(x|w)}{p_i(x)},$$

where $p_i(x)$ is the marginal distribution for expert i, that is, $p_i(x) = \int f(x|w)\pi_i(w)dw$.

The posterior for the executive is therefore

$$\begin{aligned}
\pi^*(w|x) &= \frac{\pi(w)f(x|w)}{p(x)} \\
&= \frac{\sum_{i=1}^{K} \alpha_i \pi_i(w) f(x|w)}{p(x)} \\
&= \sum_{i=1}^{K} \frac{\alpha_i p_i(x)}{p(x)} \frac{\pi_i(w) f(x|w)}{p_i(x)} \\
&= \sum_{i=1}^{K} \beta_i \pi_i^*(w|x). \qquad (10.1)
\end{aligned}$$

Here $p_i(x)$ is the marginal distribution for expert i, that is, $p_i(x) = \int f(x|w)\pi_i(w)dw$. The marginal distribution for the executive is $p(x) = \sum_{i=1}^{K} \int \alpha_i f(x|w)\pi_i(w)dw$. So, the posterior for the executive is again a weighted average of posteriors ($\pi_i^*(w|x)$, $i = 1, \cdots K$) for the K experts. The updated weight for expert i after observing $X = x$ is $\beta_i = \alpha_i p_i(x)/p(x)$. The updated weight for each expert depends on his/her marginal distribution of X. ∎

10.2 Dynamic Bayesian Ensemble Melding forecast

Forecasting ozone concentration levels in urban areas has become more and more important for public health reasons such as the safety of people like asthmatics. The forecasts can be on various time scales such as hourly or daily. The rest of this section presents how to construct a DBEM model for

hourly forecast. However, the same mode can also be used to forecast the average daily ozone concentration levels.

For the BEM model, we have hourly measurements $\hat{\boldsymbol{Z}}$ and model outputs $\tilde{\boldsymbol{Z}}_1, \cdots, \tilde{\boldsymbol{Z}}_p$ from p deterministic models, the true underlying process $\boldsymbol{Z}$ and parameters $a_1, \cdots, a_p$, $b_1, \cdots, b_p$, $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, σ_e^2, $\sigma_{\delta,1}^2, \cdots, \sigma_{\delta,p}^2$. As in DeGroot's problem, the parameter is $\boldsymbol{W} = (a_1, \cdots, a_p, b_1, \cdots, b_p, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_e^2, \sigma_{\delta,1}^2, \cdots, \sigma_{\delta,p}^2)$. The measurements $\hat{\boldsymbol{Z}}$ across space at each hour is the X in the DeGroot's problem.

We use the first K hours' measurements and model outputs to fit the BEM model, from which we obtain the posterior distributions of the parameter $\boldsymbol{W}$ for each of the K hours. These K posterior distributions will be prior distributions for $\boldsymbol{W}$ for the next hour. So we have $\pi_1^{(0)}(\boldsymbol{w}), \cdots, \pi_K^{(0)}(\boldsymbol{w})$ as in DeGroot's problem. We take the weights α_i $(i = 1, \cdots, K)$ of the prior distributions to be equal to $1/K$. Suppose we have $K+1$ hourly measurements and $K+2$ hourly model outputs. The goal is to forecast the measurement at hour $K+2$. After observing the measurements at $K+1$ hour, we obtain the updated distributions $\pi_i^{(1)}(\boldsymbol{w})$ $(i = 1, \cdots, K)$ and their corresponding weights $\alpha_i^{(1)}$. So both the K distributions for $\boldsymbol{W}$ and their weights change dynamically after the measurements become available every hour. The predictive distribution of measurements at hour $K+2$ is the following:

$$p(\hat{\boldsymbol{Z}}_{K+2}|D) = \sum_{i=1}^{K} \int \alpha_i^{(1)} f(\hat{\boldsymbol{Z}}_{K+2}|\boldsymbol{W}, \tilde{\boldsymbol{Z}}_{K+2}, \cdots, \tilde{\boldsymbol{Z}}_p) \pi_i^{(1)}(\boldsymbol{W}) d\boldsymbol{W}, \quad (10.2)$$

where D stands for all the measurements and model outputs available.

The likelihood function for $\boldsymbol{W}$ is $f(\hat{\boldsymbol{Z}}|\boldsymbol{W}, \tilde{\boldsymbol{Z}}_1, \cdots, \tilde{\boldsymbol{Z}}_p)$. To use the conclusion from DeGroot's problem, we have to derive this likelihood function. We know from that $\hat{\boldsymbol{Z}}$ and $\tilde{\boldsymbol{Z}}_1, \cdots, \tilde{\boldsymbol{Z}}_p$ have independent normal distributions given $\boldsymbol{W}$ and $\boldsymbol{Z}$. Those normal distributions are the following:

$$\begin{aligned} \hat{\boldsymbol{Z}}|(\boldsymbol{Z}, \boldsymbol{W}) &\sim \mathrm{MVN}(A_0\boldsymbol{Z}, \sigma_e^2\boldsymbol{I}) \\ \tilde{\boldsymbol{Z}}_i|(\boldsymbol{Z}, \boldsymbol{W}) &\sim \mathrm{MVN}(a_i\mathbf{1} + b_iA_i\boldsymbol{Z}, \sigma_{\delta,i}^2\boldsymbol{I}) \\ i = 1, \cdots, p. \end{aligned} \tag{10.3}$$

The matrix A_0 and $A_1, \cdots, A_p$ are the same as in the BEM model. To derive the conditional distribution $f(\hat{\boldsymbol{Z}}|\boldsymbol{W}, \tilde{\boldsymbol{Z}}_1, \cdots, \tilde{\boldsymbol{Z}}_p)$, we need to obtain the joint distribution of $f(\hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}}_1, \cdots, \tilde{\boldsymbol{Z}}_p|\boldsymbol{W})$, which is a multivariate normal distribution. For simplicity, we suppress the conditional notation $(\cdot|\boldsymbol{W})$ in the following equations.

$$
\begin{aligned}
\mathrm{E}(\hat{\boldsymbol{Z}}) &= \mathrm{E}(\mathrm{E}(\hat{\boldsymbol{Z}}|\boldsymbol{Z})) = A_0\boldsymbol{\mu} \\
\mathrm{Var}(\hat{\boldsymbol{Z}}) &= \mathrm{Var}(\mathrm{E}(\hat{\boldsymbol{Z}}|\boldsymbol{Z}))) + \mathrm{E}(\mathrm{Var}(\hat{\boldsymbol{Z}}|\boldsymbol{Z}))) \\
&= \mathrm{Var}(A_0\boldsymbol{Z}) + \sigma_e^2\boldsymbol{I} \\
&= A_0\boldsymbol{\Sigma}A_0^{\mathrm{T}} + \sigma_e^2\boldsymbol{I} \\
\mathrm{E}(\tilde{\boldsymbol{Z}}_i) &= \mathrm{E}(\mathrm{E}(\tilde{\boldsymbol{Z}}_i|\boldsymbol{Z})) = a_i\mathbf{1} + b_iA_i\boldsymbol{\mu} \\
\mathrm{Var}(\tilde{\boldsymbol{Z}}_i) &= \mathrm{Var}(\mathrm{E}(\tilde{\boldsymbol{Z}}_i|\boldsymbol{Z}))) + \mathrm{E}(\mathrm{Var}(\tilde{\boldsymbol{Z}}_i|\boldsymbol{Z}))) \\
&= \mathrm{Var}(a_i\mathbf{1} + b_iA_i\boldsymbol{Z}) + \sigma_{\delta,i}^2\boldsymbol{I} \\
&= b_i^2A_i\boldsymbol{\Sigma}A_i^{\mathrm{T}} + \sigma_{\delta,i}^2\boldsymbol{I} \\
\mathrm{Cov}(\hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}}_i) &= \mathrm{E}(\mathrm{E}(\hat{\boldsymbol{Z}}\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}|\boldsymbol{Z})) - \mathrm{E}(\hat{\boldsymbol{Z}})\mathrm{E}(\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}) \\
&= b_iA_0\boldsymbol{\Sigma}A_i^{\mathrm{T}} \\
\mathrm{Cov}(\tilde{\boldsymbol{Z}}_i, \tilde{\boldsymbol{Z}}_j) &= \mathrm{E}(\mathrm{E}(\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\tilde{\boldsymbol{Z}}_j^{\mathrm{T}}|\boldsymbol{Z})) - \mathrm{E}(\tilde{\boldsymbol{Z}}_i)\mathrm{E}(\tilde{\boldsymbol{Z}}_j^{\mathrm{T}}) \\
&= b_ib_jA_i\boldsymbol{\Sigma}A_j^{\mathrm{T}} \\
1 &\leq i < j \leq p. \qquad (10.4)
\end{aligned}
$$

Then the joint distribution of $(\hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}}_1, \cdots, \tilde{\boldsymbol{Z}}_p)$ given $\boldsymbol{W}$ is the following:

$$
\begin{pmatrix} \hat{\boldsymbol{Z}} \\ \tilde{\boldsymbol{Z}}_1 \\ \vdots \\ \tilde{\boldsymbol{Z}}_p \end{pmatrix} |\boldsymbol{W} \sim \mathrm{MVN}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}),
$$

where

$$\tilde{\boldsymbol{\mu}} = \begin{pmatrix} \tilde{\boldsymbol{\mu}}_0 \\ \tilde{\boldsymbol{\mu}}_1 \\ \vdots \\ \tilde{\boldsymbol{\mu}}_p \end{pmatrix} = \begin{pmatrix} A_0\boldsymbol{\mu} \\ a_1\mathbf{1} + b_1 A_1 \boldsymbol{\mu} \\ \vdots \\ a_p\mathbf{1} + b_p A_p \boldsymbol{\mu} \end{pmatrix}$$

and

$$\tilde{\Sigma} = \begin{pmatrix} \tilde{\Sigma}_{00} & \tilde{\Sigma}_{01} \\ \tilde{\Sigma}_{10} & \tilde{\Sigma}_{11} \end{pmatrix}.$$

We have $\tilde{\Sigma}_{00} = \mathrm{Var}(\hat{\boldsymbol{Z}})$, $\tilde{\Sigma}_{11} = \mathrm{Var}(\tilde{\boldsymbol{Z}})$ and $\tilde{\Sigma}_{01} = \mathrm{Cov}(\hat{\boldsymbol{Z}}, \tilde{\boldsymbol{Z}})$. Vector $\tilde{\boldsymbol{Z}}$ is formed by stacking $\tilde{\boldsymbol{Z}}_1, \cdots, \tilde{\boldsymbol{Z}}_p$ column by column.

Then the conditional distribution of $(\hat{\boldsymbol{Z}}|\tilde{\boldsymbol{Z}}_1, \cdots, \tilde{\boldsymbol{Z}}_p, \boldsymbol{W})$ is multivariate normal with mean

$$\tilde{\boldsymbol{\mu}}_0 + \tilde{\Sigma}_{00}\tilde{\Sigma}_{11}^{-1} \begin{pmatrix} \tilde{\boldsymbol{Z}}_1 - \tilde{\boldsymbol{\mu}}_1 \\ \vdots \\ \tilde{\boldsymbol{Z}}_p - \tilde{\boldsymbol{\mu}}_p \end{pmatrix}$$

and variance $\tilde{\Sigma}_{00} - \tilde{\Sigma}_{01}\tilde{\Sigma}_{11}^{-1}\tilde{\Sigma}_{10}$.

Unfortunately, we do not have closed forms for either β_i or $\pi_i^{(1)}(\boldsymbol{w})$ so we have to approximate them by a Monte Carlo method, so further work will be needed to implement this DBEM model. However, the Monte Carlo approximation and the implementation of this model will present some challenges.

Chapter 11

Discussion and Future Work

This chapter presents some discussion and potential future work. Section 11.1 discusses the BM model (2.9) and Section 11.2 presents some future work extending the Bayesian spatial-temporal model (5.5).

11.1 Bayesian Melding Model

The melding model proposed by Fuentes and Raftery (2005) as well as Fuentes et al. (2003) assumes the existence of a true underlying process which serves as a basis on which to "meld" the measurement and model output processes. This melding approach is a novel idea and in some sense it can deal with the fact that measurements and model output processes have different supports. However this true underlying process also presents some challenges. First, it makes the model complicated and computationally expensive. When we first consider to extend the BM model into a spatial-temporal one, we think of a approach similar to the DLM (dynamic linear model) used by West and Harrison (1997) and Huerta et al. (2004). DLM assumes the coefficients β in a linear model dynamically evolving over time.

For example, we can assume at time t, the coefficient $\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1} \sim N(\boldsymbol{\beta}_{t-1}, \boldsymbol{\Sigma})$. Because the coefficients are correlated over time, so the response is also correlated over time. However, the assumption of this true underlying process presents an obstacle to the extension of the BM model into a dynamic BM which incorporates the temporal correlation. Second, it is difficult to extend the univariate melding model into a multivariate one. For example, we cannot use the approach of Schmidt and Gelfand (2003) who propose a computationally efficient co-regionalization approach. It seems that the biggest benefit of this true underlying process is to address the differential supports issue presented by measurements and model outputs. But in fact, even meteorological and environmental work often uses some preliminary analysis to interpolate the model output onto monitoring sites and then treat the interpolates as point based data. Examples include Kasibhatla and Chameides (2000); Hogrefe et al. (2001b,a); Fiore et al. (2003, 2004). If we really can ignore the differential supports issue, then the regression approach seems more appealing and easier to implement. In that approach, the responses are the measurements from the monitoring sites and the covariate is the model output interpolated to those sites or to other locations.

Of course we have unequal numbers of model outputs and measurements, leading to a standard missing value problem. We either have missing responses or covariates.

In the melding model (2.9), we assume the mean of the true underlying process is a function of only the coordinates. However, we know that other

variables have effects on the pollutant although we do not have their information available in our application. These variables include temperature, humidity, wind speed and so on. Without the information provided by these variables, the variance of the spatial residuals $\epsilon(\boldsymbol{s})$ in model (2.9) can be large and predictions based on the true underlying process could be in question. However, when we regress measurements on the model outputs, at least part of that information is captured by in the model outputs. So the resulting residuals reflect only the part that model output fails to capture.

The regression model discussed above has the following form:

$$\hat{Z}(\boldsymbol{s}) = a(s) + b\tilde{Z}(\boldsymbol{s}) + \epsilon(s). \tag{11.1}$$

In the above model, the residuals $\epsilon(s)$ are spatially correlated with covariance matrix $\boldsymbol{\Sigma}$ constructed by correlation functions such as Matern, Gaussian or exponential. This linear model (11.1) is much easier to fit with fewer parameters and without the true underlying process. Moreover, it is not difficult to extend the model (11.1) to incorporate multivariate response and temporal correlation.

For multivariate model, we can borrow the idea from Schmidt and Gelfand (2003). They decompose the multivariate model into several independent univariate models. This decomposition enables the computation to be sequential and less time consuming. In order to extend the model (11.1) to incorporate temporal correlation, we have two potential options. The first

choice is to use DLM. The other choice is to assume the residuals $\epsilon(s)$ are correlated over time. We can assume AR or even ARMA (Auto-regressive moving average) model for the residuals over time t. If we use the dynamic linear model, the resulting spatial-temporal is easy to fit because most of the theory and results for the dynamic linear model can be applied here directly. By using the second option, the covariance matrix of residuals $\epsilon(\boldsymbol{s}, t)$ is a Kronecker product of spatial covariance matrix and the temporal covariance matrix. This complicated Kronecker product of two matrices may present some computational challenges. The above two options are both based on the assumption that the spatial and temporal correlation are separable. That is, at each time t, the spatial correlation remains the same and at each station $\boldsymbol{s}$, the temporal correlation remains the same or at least have the same structure.

11.2 Bayesian Spatial-Temporal Model

The Bayesian spatial-temporal model 5.5 has a number of limitations.

- The coefficients a, c, γ vary over space but remain constant over time. Raftery et al. (2005) also make this observation about their Bayesian averaging model and they investigate how long the historic period should be for fitting their model. For the sake of forecasting, we should use the most recent historic data but we also need a reasonably long period to have enough data to fit the model. As discussed in Section 11.1,

one possible solution is to use the DLM by allowing the coefficients to evolve in time.

- The temporal and spatial correlation may or may not be separable. Gneiting (2002) discusses a class of non-separable but stationary covariance functions for space-time data. It is possible to incorporate this class of non-separable functions into the model 5.5. Another way to model the non-separable spatial-temporal correlation is to let the autoregressive parameter ρ in model 5.5 vary over space. Similar to the coefficients a, c, γ, we can also assume ρ at different stations are spatially correlated. The autoregressive coefficients are always in the interval $(-1, 1)$, so we can use truncated normal distributions to model them. That is, for autoregressive coefficient at n locations, we have

$$(\rho_1, \cdots, \rho_n)^{\mathrm{T}} \sim c\mathrm{MVN}\boldsymbol{\mu}_\rho, \boldsymbol{\Sigma}_\rho) I[-1 < \rho_1, \cdots, \rho_n < 1],$$

  where I stands for the indicator function. That is, this indicator is 1 if all the ρ's at n locations all in the interval $(-1, 1)$, otherwise it is 0. The normalize constant c makes sure the integral of the multivariate normal density function to be 1.

- Model 5.5 assumes the spatially correlated residuals are stationary. This assumption of stationarity could be relaxed by fitting a non-stationary model, say using the celebrated Sampson-Guttorp (SAG) method of Sampson and Guttorp (1992). The main idea is that we

can first fit the two-step linear regression model (5.1), (5.2) and (5.3). Then we apply the Sampson-Guttorp method to the spatial covariance for the fitted residuals to transform the coordinates from the original geographical plane (G-plane) to the deformed plane (D-plane). After the transformation, we can use the new coordinates of the D-plane and fit the Bayesian spatial-temporal model 5.5 over again with the help of the estimated SG covariance. We can also incorporate other non-stationary process modeling approaches into the spatial-temporal model. We have already listed some reference for such approaches at the end of Section 2.1 and beginning of Section 2.2.

Bibliography

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions*, Dover.

Banerjee, S., Carlin, B. and Gelfand, A. (2003). *Hierarchical Modeling and Analysis for Spatial Data*, Chapman and Hall/CRC.

Bayarri, M. and DeGroot, M. (1989). Optimal reporting of prediction, *Journal of the American Statistical Association* **84**: 214–222.

Berliner, L. (2003). Physical-statistical modeling in geophysics, *Journal of Geophysical Research* **108**: 80–91.

Box, G., Jenkins, G. and Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control; 3rd Edition*, Holden-Day.

Brown, P. J. and Zidek, J. V. (1980). Adaptive multivariate ridge regression, *The Annals of Statistics* **8**: 64–74.

Brown, P. J. and Zidek, J. V. (1982). Multivariate regression shrinkage estimation with unknown covariance matrix, *Scandinavian Journal of Statistics* **9**: 209–215.

Brown, P. J., Le, N. D. and Zidek, J. V. (1994). Multivariate spatial interpolation and exposure to air pollutants, *The Canadian Journal of Statistics* **22**: 489–509.

Calder, C. and Cressie, N. (2006). some topics in convolution-based spatial modeling, *Proceedings of the 56th Session of the International Statistical Institute, Lisbon, Portugal, forthcoming.*

Caya, D., Laprise, R., Gigure, M., Bergeron, G., Blanchet, J., Stocks, B., Boer, G. and McFarlane, N. (1995). Description of the canadian regional climate model, *Water, Air and Soil Pollution* **82**: 477–482.

Cressie, N. A. (1993). *Statistics for Spatial Data*, Wiley-Interscience.

Cripps, E., C., C. and Kohn, R. (2003). Variable selection and covariance selection in multivariate regression models, *Technical report*, SAMSI.

DeGroot, M. (1970). *Optimal statistical decisions*, McGraw-Hill.

Edwards, A. (1967). A biometric case history: Studies on the distribution of sexes in families, 1889-1966. abstract 1286, *Biometrics.*

Fiore, A. M., Jacob, D., Liu, H., Yantosca, R. M., Fairlie, T. D. and Li, Q. (2004). Variability in surface ozone background over the united states: Implications for air quality policy, *Journal of Geophysical Research.*

Fiore, A. M., Jacob, D., Mathur, R. and Martin, R. V. (2003). Application of empirical orthogonal functions to evaluate ozone simulation with regional and global models, *Journal of Geophysical Research.*

Fuentes, M. and Raftery, A. (2002). Model validation and spatial interpolation by combining observations with outputs from numerical models via bayesian melding, *Technical Report 403*, Department of Statistics, University of Washington.

Fuentes, M. and Raftery, A. (2005). Model evaluation and spatial interpolation by bayesian combination of observations with outputs from numerical models, *Biometrics* **61**: 36–45.

Fuentes, M. and Smith, R. (2001). A new class of nonstationary spatial models, *Journal of the American Statistical Association, being reviewed.*

Fuentes, M., Guttorp, P. and Challenor, P. (2003). Statistical assessment of numerical models, *Technical Report 076*, NRCSE.

Garner, J., Lewis, T., Hogsett, W. and Andersen, C. (2005). Air quality criteria for ozone and related photochemical oxidants (second external review draft) volume iii, *Technical report*, U.S. Environmental Protection Agency.

Gelfand, A. and Smith, A. (1990). Sampling-based approaches to calculating marginal densities, *Journal of American Statistical Association* **85**: 398–409.

Gelfand, A., Schmidt, A., Banerjee, S. and Sirmans, C. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization (with discussion), *Test* **13**: 1–50.

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

Green, P. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination, *Biometrika* **82**: 711–732.

Grimit, E. P. and Mass, C. (2002). Initial results of a mesoscale short-range ensemble forecasting system over the pacific northwest, *Weather Forecasting* **17**: 192–205.

Guillas, S., Tiao, G., Wuebbles, D. and Zubrow, A. (2006). Statistical diagnostic and correction of a chemistry-transport model for the prediction of total column ozone, *Atmospheric Chemistry and Physics* **6**: 527–537.

Guttorp, P. and Walden, A. (1987). On the evaluation of geophysical models, *Geophysical Journal of the Royal Astronomical Society* **91**: 201–210.

Haas, T. (1995). Local prediction of a spatio-temporal process with an application to wet sulfate deposition, *Journal of the American Statistical Association* **90**: 1189–1199.

Haas, T. (1998). Statistical assessment of spatio-temporal pollutant trends and meteorological transport models, *Atmospheric Environment* **32**: 1865–1879.

Hastings, W. (1970). Monte carlo sampling methods using markov chains and their application., *Biometrika* **57**: 97–109.

Hawkins, D. and Cressie, N. (1984). Robust kriging-a proposal, *Journal of the International Association for Mathematical Geology* **16**: 3–18.

Higdon, D., Swall, J. and Kerv, J. (1999). Non-stationary spatial modelling, *Bayesian Statistics* pp. 761–768.

Hogrefe, C., Rao, S., Kasibhatla, P., Hao, W., Sistla, G., R., M. and McHenry, J. (2001a). Evaluating the performance of regional-scale photochemical modeling systems: Part ii-ozone predictions, *Atmos. Environ.* **35**: 4175–4188.

Hogrefe, C., Rao, S., Kasibhatla, P., Kallos, G., Tremback, C., Hao, W., Olerud, D., Xiu, A., McHenry, J. and Alapaty, K. (2001b). Evaluating the performance of regional-scale photochemical modeling systems: Part i-meteorological predictions, *Atmos. Environ.* **35**: 4159–4174.

Huerta, G., Sanso, B. and Stroud, J. (2004). A spatio-temporal model for mexico city ozone levels, *Journal of the Royal Statistical society* **53**: 231–248.

Kasibhatla, P. and Chameides, W. (2000). Seasonal modeling of regional ozone pollution in the eastern united states, *Geophys. Res. Letter* **27**: 1415–1418.

Kassteele, J. V. d., Koelemeijer, R. B. A., Dekkers, A. L. M., Schaap, M., Homan, C. and Stein, A. (2006). Statistical mapping of pm10 concentrations over western europe using secondary information from dispersion modeling and modis satellite observations, *Stochastic Environmental Research and Risk Assessment (SERRA)* **21**: 183–194.

Le, N. and Zidek, J. (1997). Interpolation with uncertain spatial covariance, *Journal of Royal Statistics Society, B* **59**: 501–510.

Le, N. D. and Zidek, J. V. (2006). *Statistical Analysis of Environmental Space-Time Process*, Springer.

Lindley, D. and Smith, A. (1972). Bayes estimate for the linear model, *Journal of the Royal Statistical Society. Series B (Methodological).*

Liu, Z. (2007). Bayesian melding model software in r, in preparation, *Technical report*, Department of Statistics, The University of British Columbia.

Mardia, K. V. and Goodall, C. R. (1993). *Multivariate Environmental Statistics: Spatial-temporal analysis of multivariate environmental monitoring data*, Elsevier Science Publishers.

Matern, B. (1960). Spatial variation, *Technical Report 49*, Meddelanden fran Statens Skogsforskningsinstitut.

Matheron, G. (1962). Trait de geostatistique appliquee, tome i, *Memoires du Bureau de Recherches Geologiques et Minieres.*

Matheron, G. (1963). Traite de geostatistique appliquee, *Technical report*, Memoires du Bureau de Recherches Geologiques et Minieres.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953). Equation of state calculations by fast computing machines., *Journal of Chemical Physics* **21**: 1087–1092.

Nychka, D., Wikle, C. and Royle, J. (2002). Multiresolution models for nonstationary spatial covariance functions, *Statistical Modelling: An International Journal* **2**: 315–331.

Odman, M. and Ingram, C. (1996). Multiscale air quality simulation platform (maqsip) source code documentation and validation, *Technical report*, Environmental Programs, MCNC-North Carolina Supercomputing Center.

Paciorek, C. and Schervish, M. (2004). Nonstationary covariance functions for gaussian process regression, *in* S. Thrun, L. Saul and B. Schölkopf (eds), *Advances in Neural Information Processing Systems 16*, MIT Press, Cambridge, MA.

Paciorek, C. and Schervish, M. (2006). Spatial modelling using a new class of nonstationary covariance functions, *Environmetrics* **17**: 483–506.

Poole, D. and Raftery, A. (2000). Inference for deterministic simulation models: The bayesian melding approach, *Journal of American Statistical Association* **95**: 1244–1255.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Raftery, A. E., Gneiting, T., Balabdaoui, F. and Polakowski, M. (2005). Using bayesian model averaging to calibrate forecast ensembles, *Monthly Weather Review* **133**: 1155–1174.

Ramsay, J., Hooker, G., Campbell, D. and Cao, J. (2007). Parameter estimation for differential equations: a generalized smoothing approach, *Journal of the Royal Statistical Society: Series B* **69**: 741C796.

Ribeiro, P. J. and Diggle, P. (2001). geor: a package for geostatistical analysis, *R-NEWS* **1**(2): 14–18. ISSN 1609-3631.

Sacks, J., Welch, W., Mitchell, T. and Wynn, H. (1989). Design and analysis of computer experiments, *Statistical Science* **4**: 409–435.

Sampson, P. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure, *Journal of the American Statistical Association* **87**: 108–119.

Sanso, B. and Guenni, L. (2002). Combining observed rainfall and deterministic prediction using a bayesian approach, *Technical report*, CESMA.

Schmidt, A. M. and Gelfand, A. E. (2003). A bayesian coregionalization approach for multivariate pollutant data, *ournal of Geophysical Research-Atmospheres.*

Smith, A. (1973). A general bayesian linear model, *Journal of the Royal Statistical Society. Series B (Methodological).*

Speed, T. (1983). Modelling, *CSIRO DMS Newsletter* **96**: 1–2.

Stein, M. (1999). *Interpolation of spatial data : some theory for kriging,* New York : Springer.

Sun, W., Le, N. and Zidek, J. (1998). Assessment of bayesian multivariate interpolation approach for health impact studies, *Environmetrics* **9**: 565–586.

van der Merwe, A. and Zidek, J. V. (1980). Multivariate regression analysis and canonical variates, *The Canadian Journal of Statistics* **8**: 27–39.

Wackernagel, H. (1998). *Multivariate Geostatistics - An Introduction with Applications, 2nd Edition*, Springer-Verlag.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models; 2 edition*, Springer.

Wheeler, N. and Houyoux, M. (1998). Development and implementation of a seasonal model for regional air quality, *Proc. Air and Waste Manage Association.*

Wikle, C. (2001). A kernel-based spectral approach for spatio-temporal dynamic models, *Proceedings of the 1st Spanish Workshop on Spatio-Temporal Modelling of Environmental Processes (METMA)* pp. 167–180.

Wikle, C. and Berliner, L. (2005). Combining information across spatial scales, *Technometrics* **47**: 80–91.

Wikle, C. and Cressie, N. (1999). A dimension reduced approach to space-time kalman filtering, *Biometrika* **86**: 815–829.

Wikle, C. and Royle, J. (1999). Space-time models and dynamic design of environmental monitoring networks, *Journal of Agricultural, Biological, and Environmental Statistics* **4**: 489–507.

Wikle, C., Milliff, R., Nychka, D. and Berliner, L. (2001). Spatio-temporal hierarchical bayesian modeling: Tropical ocean surface winds, *Journal of American Statistical Association* **96**: 382–397.

Yaglom, A. (1987). *Correlation Theory of Stationary and Related Random Functions*, New York: Springer-Verlag.

Zidek, J., Sun, W. and Le, N. (2000). Designing and integrating composite networks for monitored multivariate gaussian pollution fields, *Journal of the Royal Statistical society* **49**: 63–79.